



Inactivation des centromères et élimination programmée d'ADN chez le cilié *Paramecium tetraurelia*

Maoussi Lhuillier-Akakpo

► To cite this version:

Maoussi Lhuillier-Akakpo. Inactivation des centromères et élimination programmée d'ADN chez le cilié *Paramecium tetraurelia*. Génétique. Université Pierre et Marie Curie - Paris VI, 2014. Français. NNT : 2014PA066160 . tel-01309804

HAL Id: tel-01309804

<https://theses.hal.science/tel-01309804>

Submitted on 1 May 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THESE DE DOCTORAT
DE L'UNIVERSITÉ PIERRE ET MARIE CURIE**

Spécialité Génétique
Ecole Doctorale : Complexité du Vivant

Présentée par **Maoussi Lhuillier-Akakpo**

**Inactivation des centromères et élimination programmée
d'ADN chez le cilié *Paramecium tetraurelia***

Soutenue le 29 avril 2014

Devant le jury composé de :

Mme Emmanuelle d'Alençon

M. Slimane Ait-Si-Ali

Mme Valérie Borde

M. Raphaël Margueron

Membre invité : Mme Sandra Duharcourt

Rapporteur

Rapporteur

Examineur

Examineur

Directrice de thèse

“Si la science évolue, c'est souvent parce qu'un aspect encore inconnu des choses se dévoile soudain”

François Jacob

REMERCIEMENTS

Je tiens d'abord à remercier tous les membres de mon jury pour avoir accepté d'évaluer mon travail et je remercie particulièrement mes deux rapporteurs pour le temps qu'ils ont pris à lire ma thèse.

Ensuite, un grand merci à Sandra Duharcourt, pour son encadrement pendant ces quatre années passées au laboratoire, pour sa pédagogie, ses qualités scientifiques et son dévouement qui m'ont permis d'avancer et de progresser scientifiquement. Merci de m'avoir donné la chance de mettre en valeur mon travail en m'encourageant notamment à participer à de nombreux congrès et en me faisant participer à l'écriture d'une revue scientifique.

Merci à tous les membres présents et passés du laboratoire : Andrea, Augustin, Frédéric, Mélody, Sandrine, Véronique, et à nos deux stagiaires Amandine et Thomas qui ont pris la relève ! L'ambiance du labo a toujours été joyeuse et stimulante scientifiquement, je les remercie pour ça.

Mon aventure « paramécie » a commencé à l'ENS, pour un stage de M2 sous la direction de Sandra, avant que nous ne déménagions à l'Institut Jacques Monod. Alors, merci à l'équipe d'Eric Meyer avec qui j'ai travaillé pendant plus d'un an : Deepankar, Baptiste, Sophie, Simone, Simran, Gérard, Quentin, Guillaume. Et je n'oublierai pas les autres équipes « paramécies » : Linda, Cyril, Olivier pour leur travail monstrueux sur toutes les données de séquençage et l'équipe de Mireille Bétermier avec qui j'ai pu interagir de nombreuses fois : Julien, Antoine, Emeline, Vinciane et ceux que j'oublie !

Je remercie également les membres de mon comité de thèse : Raphaël Margueron, Frédérique Peronnet, Eric Meyer et Catherine Alcaïde pour leurs conseils et leurs avis critiques sur mon travail au cours de ma thèse.

Merci beaucoup à tous mes amis de l'IJM et d'ailleurs, qui ont vécu cette thèse avec moi : Anne-Laure, Fany, Aurélia, Aurélie, Lénaïg, Kévin, Vincent, Marie, Jeanne, Julien, Julie, Anaïs... pardon si j'en oublie. Parce qu'il est toujours bon de se détendre autour d'un café (ou d'un thé pour moi) et de partager ses (més)aventures de thèse !

* Et spécial merci à celui qui est devenu un peu plus qu'un ami et qui m'a soutenu pendant toute la fin de ma thèse *

Un grand merci à la plate-forme d'imagerie de l'institut qui nous a bien aidés dans la production de toutes nos images.

Je tiens également à remercier toutes les personnes qui nous aident à travailler au quotidien, Stéphane notre gestionnaire, Norry, gérant du magasin et Patricia à la laverie.

Je remercie bien sûr ma famille, qui a toujours soutenu mes choix et m'a permis d'en être ici.

Enfin, merci à tous ceux qui ont apporté une aide, scientifique ou non, à l'écriture de ce manuscrit.

Ah oui, et merci à Coca Cola, Lipton, (parce qu'on peut écrire une thèse sans boire de café), aux madeleines St-Michel du distributeur du 2^{ème} étage du bâtiment A (parce que j'ai toujours un petit creux au labo) et aux JO d'hiver qui m'ont apporté de bons moments de détente pendant l'écriture !

Inactivation des centromères et élimination programmée d'ADN chez le cilié

Paramecium tetraurelia

RESUME :

Chez les eucaryotes, les génomes sont organisés en différents domaines chromatinien qui facilitent l'accomplissement de nombreuses fonctions nucléaires comme la transcription, la réparation de l'ADN ou la ségrégation des chromosomes. Le cilié *Paramecium tetraurelia* est utilisé comme organisme modèle pour étudier les mécanismes généraux qui maintiennent l'intégrité des génomes eucaryotes. Chez cet organisme unicellulaire, deux micronoyaux, diploïdes, assurent les fonctions germinales en permettant la transmission du matériel génétique à la génération sexuelle suivante tandis qu'un macronoyau, hautement polyploïde, assure les fonctions somatiques en permettant la transcription des gènes. A chaque cycle de reproduction sexuée, le macronoyau est perdu et un nouveau macronoyau est produit à partir du noyau zygotique. Le développement macronucléaire est associé à des événements d'élimination massifs et reproductibles d'ADN micronucléaire. D'une part, des éléments répétés (transposons, régions minisatellites), de plusieurs kilobases de long sont imprécisément éliminés, conduisant à la fragmentation des chromosomes macronucléaires. D'autre part, des séquences courtes et uniques, appelées IES, sont éliminées très précisément au nucléotide près. Les mécanismes qui régulent l'élimination programmée d'ADN au cours du développement macronucléaire semblent similaires à la mise en place de structures chromatinien répressives chez les eucaryotes supérieurs et impliquent notamment une voie de l'ARN interférence spécifique du cycle sexuel. Nous avons montré que la triméthylation de l'histone H3 sur la lysine 9 et la lysine 27 (H3K9me3 et H3K27me3) apparaît transitoirement dans le macronoyau en développement au moment où se produisent les événements d'élimination d'ADN. Nous avons identifié une histone méthyltransférase putative, Ezl1p, requise pour la déposition de ces marques d'histones et qui s'avère être nécessaire pour l'élimination programmée d'ADN. De manière intéressante, seul un sous-ensemble de séquences dépend d'EZL1 pour être éliminé, les séquences les plus courtes étant souvent insensibles à l'histone méthyltransférase. De la même manière, les petits ARN non codants produits pendant le cycle sexuel ne sont requis que pour l'élimination d'un sous-ensemble des séquences sensibles à EZL1. Ces résultats suggèrent que plusieurs voies, partiellement chevauchantes, sont impliquées dans l'élimination programmée d'ADN chez *P. tetraurelia*.

Le maintien de l'intégrité des génomes repose également sur les régions centromériques qui, à chaque division cellulaire, assurent la ségrégation fidèle des chromosomes. Chez tous les eucaryotes, les centromères fonctionnels sont définis par la présence d'un variant d'histone H3, appelé CenH3 qui remplace l'histone canonique H3 dans certains nucléosomes. Cependant, le contrôle de l'identité des centromères dans les cellules eucaryotes est mal compris. Le cilié *Paramecium tetraurelia* s'avère également être un modèle de choix pour comprendre les mécanismes qui régulent l'activité des centromères. Au cours de la croissance végétative, les micronoyaux se divisent par mitoses classiques tandis que le macronoyau se divise par un processus non mitotique qui n'impliquerait ni condensation des chromosomes ni formation de fuseau mitotique. Le macronoyau se comporte donc comme s'il était dépourvu de fonction centromérique. Nous avons identifié le variant centromérique CenH3 de *P. tetraurelia* et mis en évidence que la fonction centromérique du macronoyau est perdue au cours du développement macronucléaire, à un moment qui coïncide avec les événements d'élimination d'ADN. Nous avons montré que l'endonucléase putative PiggyMac, requise pour l'élimination de toutes les séquences d'ADN pendant le développement, ainsi que l'histone méthyltransférase putative Ezl1p sont nécessaires à l'inactivation des centromères dans le macronoyau en développement. Ces résultats suggèrent que les centromères sont inactivés par l'élimination physique des séquences centromériques et ouvrent des perspectives pour comprendre les mécanismes qui régulent l'activité des centromères chez les eucaryotes.

TABLE DES MATIERES

PREAMBULE.....	1
INTRODUCTION.....	3
CHAPITRE 1. REPRESSION CHROMATINIENNE ET MAINTIEN DE L'INTEGRITE DES GENOMES.....	4
I. Les différents types de chromatine répressive.....	4
1. L'hétérochromatine.....	4
2. Répression par les complexes Polycomb.....	5
3. La chromatine « noire ».....	5
II. Complexes protéiques impliqués dans la mise en place des structures chromati- niques répressives.....	7
1. Les acteurs de l'hétérochromatine.....	7
2. Les complexes Polycomb.....	9
III. Ciblage des facteurs à leurs séquences cibles.....	11
1. Recrutement par des ARN non codant.....	11
2. Recrutement par des séquences d'ADN ou des facteurs de fixation à l'ADN.....	13
IV. Assemblage des structures chromatiennes répressives et inhibition de l'activité transcriptionnelle des séquences cibles.....	15
1. Condensation de l'hétérochromatine.....	15
2. Répression transcriptionnelle par les protéines PcG.....	17
V. Formation de structures nucléaires d'ordre supérieur.....	19
1. Cluster hétérochromatiques.....	19
2. Polycomb bodies.....	20
CHAPITRE 2. LES CENTROMERES, ACTEURS DU MAINTIEN DE L'INTEGRITE DES GENOMES.....	21
I. Le centromère, structure générale.....	23
II. Déterminants de l'identité centromérique.....	25
1. L'ADN centromérique.....	25
2. Cenp-A/CenH3 : marqueur des centromères actifs.....	27
III. Régulation épigénétique de l'activité des centromères, un secret bien gardé... ..	30
1. Le rôle de CenH3 dans son propre recrutement aux centromères.....	30
2. Rôle des facteurs d'interaction avec CenH3.....	31
3. Rôle de la chromatine centromérique.....	33
4. Un rôle pour la transcription ?	34

IV. Centromères : quand stabilité rime avec plasticité.....	35
1. Chromosomes dicentriques.....	35
2. Stabilisation des chromosomes dicentriques par inactivation des centromères.....	37
CHAPITRE 3. <i>PARAMECIUM TETRAURELIA</i> : Organisme modèle pour l'étude de la	
régulation épigénétique de l'organisation du génome.....	43
I. Dimorphisme nucléaire et cycle de vie.....	43
II. Génome...s	45
III. Réarrangements du génome associés au développement macronucléaire.....	47
1. Endoréplication.....	47
2. Elimination imprécise de séquences répétées.....	47
3. Elimination précise des IES.....	49
4. L'endonucléase domestiquée PiggyMac.....	51
5. Réparation des cassures double brin.....	53
IV. Ciblage de la machinerie d'excision aux régions éliminées.....	53
1. Régulation épigénétique des réarrangements du génome.....	55
2. ARN non codants et réarrangements programmés du génome.....	56
3. Le modèle de « genome scanning ».....	59
4. Ciblage de l'endonucléase par les scanARN au niveau des séquences à éliminer.....	61
OBJECTIF DE THESE N°1.....	63
V. Dimorphisme nucléaire et division cellulaire.....	64
OBJECTIF DE THESE N°2.....	65
RESULTATS.....	67
PARTIE 1. Rôle de la méthylation des histones dans les réarrangements programmés du génome	
chez <i>Paramecium tetraurelia</i>.....	69
Article 1 (en préparation)	
Résultats complémentaires : Analyse à l'échelle du génome entier de la rétention des IES dans les	
cellules inactivées pour PTIWI01 et PTIWI09.....	71
PARTIE 2. Mécanismes d'inactivation programmée des centromères au cours du développement	
macronucléaire de <i>Paramecium tetraurelia</i>.....	75
Article 2 (en préparation)	
Résultats complémentaires : Identification de séquences centromériques par immunoprécipitation	
de la chromatine (ChIP).....	77
CONTRIBUTIONS ADDITIONNELLES.....	97

DISCUSSION.....	99
I. Ezl1p, une protéine Polycomb actrice de l'élimination programmée d'ADN chez <i>Paramecium</i>	101
1. Méthylation des histones et élimination imprécise des régions mic-spécifiques.....	101
2. Méthylation des histones et excision des IES.....	105
II. Des voies partiellement chevauchantes impliquées dans l'élimination d'ADN chez <i>Paramecium</i> : complexification du modèle de scanning.....	108
1. EZL1 affecte un sous-ensemble des IES sensibles à PGM.....	108
2. DCL2 et DCL3 affectent un sous-ensemble des IES sensibles à EZL1.....	111
III. Rôles alternatifs de EZL1 et des scanARN.....	113
IV. Régulation de l'activité centromérique chez <i>Paramecium tetraurelia</i>	117
1. Les centromères : des séquences mic-spécifiques ?	117
2. Mécanismes de l'inactivation des centromères.....	119
REFERENCES.....	123
ANNEXES.....	141
Annexe 1 : Genome-defence small RNAs exapted for epigenetic mating-type inheritance (<i>Singh et al., accepté dans Nature</i>).....	141
Annexe 2 : RNA-guided DNA rearrangements in ciliates : is the best genome defense a good offense ? <i>Biol Cell</i> . 2012 Jun;104(6):309-25.....	143

LISTE DES FIGURES

Figure 1. Les différents types de chromatine répressive.....	6
Figure 2. Complexes protéiques impliqués dans la mise en place des différents types de chromatine répressive.....	8
Figure 3. Recrutement des facteurs de l'hétérochromatine et des PcG par des ARN non codants.....	10
Figure 4. Mécanismes de répression par HP1 et par le complexe PRC1.....	16
Figure 5. Représentation schématique des «Polycomb bodies» dans le noyau.....	18
Figure 6. Structure générale des centromères.....	22
Figure 7. ADN centromériques de différents eucaryotes.....	24
Figure 8. Structure des protéines CenH3.....	28
Figure 9. Modèle de la structure dynamique des nucléosomes centromériques au cours du cycle cellulaire.....	32
Figure 10. Mécanismes d'inactivation des centromères.....	36
Figure 11. Phylogénie des eucaryotes.....	42
Figure 12. Cycle de vie de <i>Paramecium tetraurelia</i>	44
Figure 13. Elimination imprécise de séquences répétées au cours du développement macronucléaire de <i>P. tetraurelia</i>	46
Figure 14. Elimination précise des IES.....	48
Figure 15. Mécanismes moléculaires impliqués dans l'excision des IES chez <i>Paramecium</i>	52
Figure 16. Hérité mendélienne et non mendélienne.....	54
Figure 17. Le modèle de « genome scanning » chez <i>Paramecium</i>	58
Figure 18. Distribution des scores de rétention des IES dans le génome des cellules inactivées pour PTIW101 et PTIW109.....	72
Figure 19. Les étapes du ChIP.....	76
Figure 20. Mise au point du ChIP-1.....	82
Figure 21. Mise au point du ChIP-2.....	86
Figure 22. Résultats du ChIP CenH3.....	90
Figure 23. Modèle du mécanisme d'élimination imprécis des séquences micronucléaires.....	104

Figure 24. EZL1 et excision des IES.....	106
Figure 25. Les différentes voies impliquées dans l'élimination des IES chez <i>Paramecium</i>	110
Figure 26. Modèle expliquant la formation de foci EZL1, H3K27me3 et H3K9me3 au cours du développement macronucléaire.....	114
Figure 27. Modèle du mécanisme d'inactivation des centromères dans le macronoyau en développement chez <i>Paramecium</i>	120

PREAMBULE

Dans les cellules eucaryotes, l'ADN est assemblé avec des protéines histones et non histones pour former la chromatine. Chaque unité de chromatine, ou nucléosome, contient 145 à 147 paires de bases (pb) d'ADN enroulé autour d'un octamère d'histones. Les nucléosomes dits canoniques sont constitués d'une paire de chaque histone H3, H4, H2A et H2B. Dans les génomes, la chromatine est façonnée par des enzymes qui modifient les histones, modifient l'ADN ou encore permettent l'incorporation de variants d'histones pour former des domaines qui facilitent l'organisation de nombreuses fonctions nucléaires comme la transcription, la recombinaison, la réparation de l'ADN ou la ségrégation des chromosomes (Erdel et al., 2013; Sarma and Reinberg, 2005). Comprendre les mécanismes qui permettent le ciblage, la mise en place et le maintien de ces domaines chromatinien est essentiel pour comprendre comment se maintient l'intégrité des génomes eucaryotes.

Au laboratoire, nous utilisons le cilié *Paramecium tetraurelia* comme organisme modèle pour étudier les mécanismes épigénétiques qui contrôlent l'organisation et la stabilité des génomes eucaryotes. Chez cet organisme unicellulaire, les fonctions somatiques et germinales sont séparées dans deux types de noyaux distincts qui partagent le même cytoplasme. Le génome somatique dérive du génome germinale après un processus de différenciation qui est associé à l'élimination massive et reproductible de séquences d'ADN répétées et de séquences uniques dérivées de transposons. La reconnaissance de ces séquences fait intervenir des mécanismes épigénétiques qui semblent similaires aux systèmes de répression transcriptionnelle médiée par la chromatine chez les autres eucaryotes.

Le noyau somatique et le noyau germinale se distinguent également par leurs activités centromériques. Les centromères sont des régions chromosomiques permettant la ségrégation fidèle des chromosomes et sont donc essentielles à l'intégrité des génomes eucaryotes. Ils sont spécifiés par une chromatine particulière contenant un variant d'histone H3 centromérique (Allshire and Karpen, 2008). Chez *Paramecium*, le noyau somatique serait dépourvu de centromères actifs impliquant qu'il existe un mécanisme d'inactivation des centromères germinaux pendant la différenciation du génome somatique. L'étude des processus impliqués

dans l'inactivation des centromères chez *Paramecium* est informative sur les mécanismes qui contrôlent l'identité des centromères et régulent leur activité chez les eucaryotes.

Une vaste réorganisation chromatinienne intervient pendant la différenciation du noyau somatique de *Paramecium*. Mon travail de thèse consisté à :

1. Etudier le rôle des modifications post-traductionnelles d'histones dans les mécanismes d'élimination d'ADN pendant la différenciation du génome somatique. Nous avons mis en évidence que des marques d'histones connues pour être associées à des domaines chromatinien qui répriment l'activité transcriptionnelle sont impliquées.
2. Caractériser les centromères de *Paramecium tetraurelia* et étudier les mécanismes programmés qui conduisent à leur inactivation pendant la différenciation somatique.

Dans l'introduction qui va suivre, je présenterai donc dans un premier temps les mécanismes connus de répression transcriptionnelle médiée par la chromatine chez les eucaryotes. Je détaillerai ensuite l'état des connaissances sur les mécanismes moléculaires qui régulent l'activité des centromères. Enfin, j'expliquerai en quoi le cilié *Paramecium tetraurelia* est un organisme modèle particulièrement adapté pour l'étude des processus qui régulent la mise en place de ces domaines chromatinien et maintiennent la stabilité des génomes.

INTRODUCTION

CHAPITRE 1

REPRESSION CHROMATINIENNE ET MAINTIEN DE L'INTEGRITE DES GENOMES

Les régions non codantes des génomes eucaryotes sont parsemées de séquences répétées et d'éléments transposables potentiellement mutagènes, dont l'activité doit être maintenue à un niveau basal. Les régions codantes contiennent quant à elles des gènes qui doivent être différenciellement exprimés pour assurer les bonnes fonctions cellulaires au bon moment. Pour maintenir l'intégrité de leur génome et la stabilité de leur identité, les cellules eucaryotes ont développé des systèmes qui visent à limiter l'activité transcriptionnelle des régions codantes et non codantes du génome. Ces mécanismes sont généralement basés sur la compaction des nucléosomes qui composent la chromatine, réduisant l'accessibilité des machineries enzymatiques au niveau des séquences concernées.

I. Les différents types de chromatine répressive

1. L'hétérochromatine

La formation d'hétérochromatine est un exemple classique de mécanisme permettant de réprimer certaines régions du génome. Cytologiquement, l'hétérochromatine est observable dans les noyaux interphasiques et correspond à une structure particulièrement dense aux électrons, traduisant une compaction importante des nucléosomes. Elle est historiquement opposée à l'euchromatine, la fraction décondensée du génome (Grewal and Jia, 2007). D'un point de vue moléculaire, l'hétérochromatine est caractérisée par l'hypoacétylation des queues d'histones, par la di- ou la triméthylation de la lysine 9 de l'histone H3 (H3K9me2 ou me3) et par la présence du facteur hétérochromatique HP1 (Lachner et al., 2001; Litt et al., 2001; Nakayama et al., 2001). Les régions chromosomiques contenant de nombreuses séquences répétées telles que les régions péricentromériques, riches en ADN satellite et en transposons, et les télomères sont les principales cibles de l'hétérochromatine. Sa topologie compactée limite l'activité transcriptionnelle des

éléments transposables potentiellement mutagènes et prévient la recombinaison entre des motifs de séquences répétés (Birchler et al., 2000). L'hétérochromatine participe également à la structuration des chromosomes, elle intervient notamment dans la cohésion des chromatides sœurs en mitose en recrutant la cohésine (Bernard et al., 2001). Enfin, des gènes spécifiques présents dans des régions euchromatiques peuvent être ciblés par des facteurs de l'hétérochromatine pour être régulés (Yasuhara and Wakimoto, 2006).

2. Répression par les complexes Polycomb

Si l'hétérochromatine assure essentiellement la répression des séquences répétées, un certain nombre de gènes, impliqués dans le développement, la différenciation ou la régulation du cycle cellulaire, doivent également être stablement réprimés pour déterminer et maintenir l'identité des cellules. La répression transcriptionnelle durable et héritable de gènes est généralement dépendante des protéines Polycomb (PcG), un groupe de facteurs historiquement associé à la répression des gènes homéotiques chez la Drosophile. Les protéines PcG sont maintenant connues pour réguler l'expression de plusieurs centaines de gènes impliqués dans de nombreuses fonctions cellulaires. La chromatine associée aux complexes PcG est caractérisée par la triméthylation de la lysine 27 de l'histone H3 (H3K27me3) qui intervient, entre autre, dans un processus de condensation des nucléosomes (Schwartz and Pirrotta, 2013; Simon and Kingston, 2013). Dans la littérature, ce type de chromatine est parfois qualifié d'hétérochromatine facultative ou parfois distingué de l'hétérochromatine pour être simplement décrit comme « chromatine Polycomb ». Par souci de clarté, seules les régions associées à H3K9me2/me3 et HP1 seront qualifiées d'hétérochromatine dans la suite de cette introduction.

3. La chromatine « noire »

Chez la Drosophile, la cartographie par CHIP des sites de fixation à l'ADN de 53 protéines a mis en évidence un troisième type de chromatine répressive, appelée « chromatine noire » (BLACK chromatin). Cette chromatine couvre 48% du génome des cellules embryonnaires de Drosophile et forme de larges domaines de plus de 100 kb qui contiennent de nombreux gènes régulés au cours du développement. La chromatine noire est caractérisée par la présence de la protéine LAM, composante de la lamina nucléaire, suggérant que cette chromatine est localisée à la périphérie du

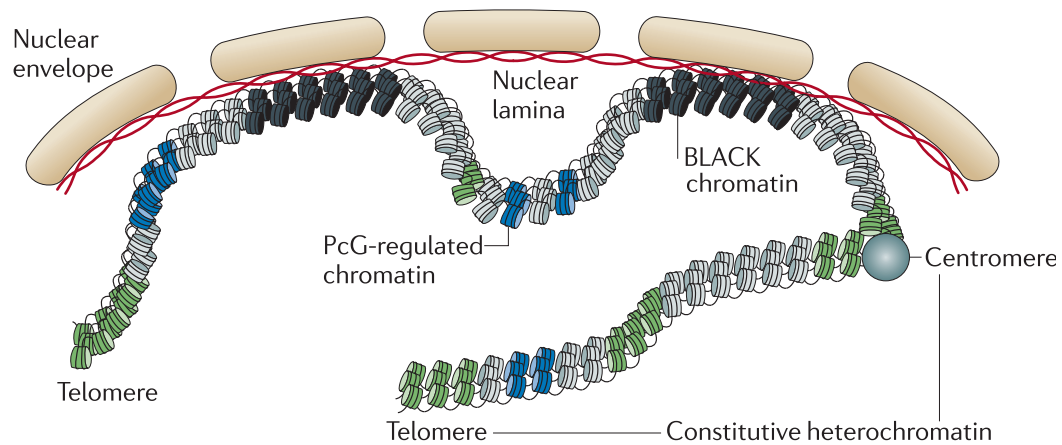


Figure 1. Les différents types de chromatine répressive. L'hétérochromatine, en vert, est associée constitutivement aux centromères et aux télomères mais peut également être retrouvée plus localement dans des régions géniques. La chromatine associée aux protéines Polycomb est représentée en bleu. L'hétérochromatine «noire» est associée à l'enveloppe nucléaire. Sa fonction et sa composition ne sont pas encore caractérisées.

Adapté de Beisel and Paro, Nat. Rev., 2011

noyau et interagit avec l'enveloppe nucléaire (Filion et al., 2010). En plus de réprimer activement certains gènes, la chromatine noire pourrait également jouer un rôle structural et favoriser l'organisation tridimensionnelle du génome au sein du noyau. Les modifications d'histones qui pourraient être associées à la chromatine noire sont, pour l'instant, inconnues. Cependant, l'association de la chromatine noire avec EFF, une enzyme de conjugaison à l'ubiquitine, suggère que celle-ci contient des marques spécifiques d'ubiquitylation.

La Figure 1 présente les différents types de chromatine répressive connus et leur localisation nucléaire. Mis à part le cas de la chromatine noire dont l'identification est très récente, les facteurs et les mécanismes impliqués dans la mise en place de la chromatine répressive sont de mieux en mieux caractérisés. La formation d'hétérochromatine et la régulation des gènes par les Polycomb présentent d'ailleurs de nombreux points communs, à commencer par le rôle crucial des modifications d'histone dans leur mise en place et l'implication de facteurs reconnaissant ces modifications pour induire une répression transcriptionnelle.

II. Complexes protéiques impliqués dans la mise en place des structures chromatinienne répressives

1. Les acteurs de l'hétérochromatine

Chez *Drosophila melanogaster*, la levure *Schizosaccharomyces pombe*, la plante *Arabidopsis thaliana* ou encore les mammifères, les lysine méthyltransférases de la famille SU(VAR)3-9 catalysent la di- ou la triméthylation de H3K9 dans les régions péri-centromériques, les régions télomériques, au niveau des transposons ou dans d'autres sites euchromatiques (Jackson et al., 2004; Rea et al., 2000; Rice et al., 2003; Schotta, 2002). L'activité catalytique des protéines SU(VAR)3-9 est détenue par leur domaine SET situé dans la région C-terminale. Ce domaine est également retrouvé chez les protéines Polycomb (SET = SUVAR, E(z), TRITORAX) (Schotta, 2002). La méthylation de H3K9 établit un site de fixation pour les protéines HP1. Identifiée en premier lieu chez la Drosophile, HP1 appartient à une famille de protéines chromatinienne dont des représentants sont retrouvés dans de nombreux organismes, depuis la levure *S. pombe* (Swi6), jusqu'à l'Homme qui en possède trois isoformes (HP1 α , HP1 β et HP1 γ). Ces protéines possèdent un chromodomaine qui reconnaît spécifiquement les marques de méthylation

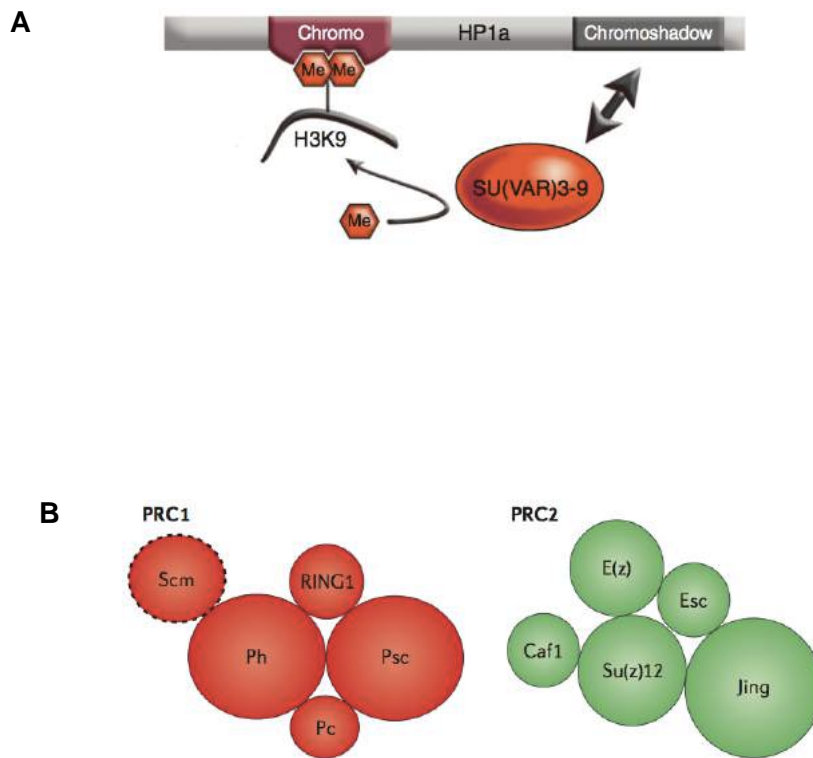


Figure 2. Complexes protéiques impliqués dans la mise en place des différents types de chromatine répressive. **A.** Dans le cas de l'hétérochromatine, la lysine méthyltransférase SU (VAR)3-9 catalyse la di- ou la triméthylation de H3K9. HP1 interagit avec H3K9me2/3 par son chromodomaine et avec SU(VAR)3-9 par son chromoshadow domaine. **B.** Les complexes Polycomb canoniques chez *Drosophila melanogaster*. Le complexe PRC1 est composé des protéines Polyhoméotic (Ph), Posterior sex combs (Psc), Polycomb (Pc) et RING1. Le facteur Sex comb on midleg (Scm) interagit moins fortement avec le reste du complexe. Pc contient un chromodomaine qui fixe la triméthylation de H3K27. Le complexe PRC2 comprend la lysine méthyltransférase E(z) responsable de la méthylation de H3K27 ; E(z) s'associe à Extra sexcomb (Esc) et à Su(z)12. L'histone chaperone Caf1 et la protéine à doigt de zinc Jing font également partie du complexe.

Elgin and Reuter, Cold Spring Harb Perspect Biol., 2013

Schwartz and Pirrota, Nat. Rev., 2013

H3K9me2/me3 (Jacobs et al., 2001). Le chromodomaine est suivi par un domaine charnière flexible, de séquence et de longueur variable et d'un chromoshadow domaine qui participe à la dimérisation de HP1 et au recrutement d'une variété d'autres facteurs de modification de la chromatine (Lomberk et al., 2006). HP1 interagit notamment avec SU(VAR)3-9, ce qui stabilise sa fixation à H3K9me2/me3 (Eskeland et al., 2006) (Figure 2A).

2. Les complexes Polycomb

Les protéines du groupe Polycomb (PcG) ont été découvertes il y a plusieurs dizaines d'années chez *Drosophila melanogaster* comme des répresseurs des gènes homéotiques pendant le développement embryonnaire. En plus des insectes, la fonction des protéines Polycomb est maintenant démontrée chez les plantes et les vertébrés (Schwartz and Pirrotta, 2013).

Deux complexes multiprotéiques majeurs, PRC1 et PRC2 (PRC = Polycomb Repressive Complex) sont impliqués dans la répression transcriptionnelle par les protéines Polycomb (Figure 2B). Chez la Drosophile, le complexe PRC2 contient la lysine méthyltransférase Enhancer of zeste (E(z)), capable de di- ou trimétyler la lysine 27 de l'histone H3 (H3K27me2/me3) grâce à son domaine SET catalytique (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002). E(z) s'associe à Extra sexcombs (Esc) et à Su(z)12. Ces deux facteurs sont essentiels à l'action de E(z) car ils interagissent avec les nucléosomes de la région cible et régulent l'activité de la lysine méthyltransférase (Cao and Zhang, 2004; Ketel et al., 2005). La protéine chaperone Caf1 (aussi connue sous le nom de Nurf55) fixe Su(z)12 et contribue à l'activité du complexe PRC2 tandis que la protéine à doigt de zinc Jing participe à la stabilité du complexe, au moins au niveau de certains gènes cibles (Cao and Zhang, 2004; Kim et al., 2009). Chez les mammifères, les éléments du complexe PRC2 possèdent leurs homologues directs. Deux gènes, EZH1 et EZH2 codent pour des homologues de E(z) qui interagissent avec EED (embryonic ectoderm development), l'homologue de Esc, et SUZ12. Les protéines RBBP4 et RBBP7 sont des homologues de Caf1 (Di Croce and Helin, 2013; Schwartz and Pirrotta, 2013). Il est à noter que les lysine méthyltransférases impliquées dans la méthylation des histones possèdent également des cibles protéiques non-histones. Dans des cellules humaines, EZH2 est notamment capable de monométyler le récepteur nucléaire ROR α sur la lysine K38 conduisant à son ubiquitylation et à sa dégradation (Lee et al., 2012). EZH2 participe donc à la régulation de l'expression des gènes

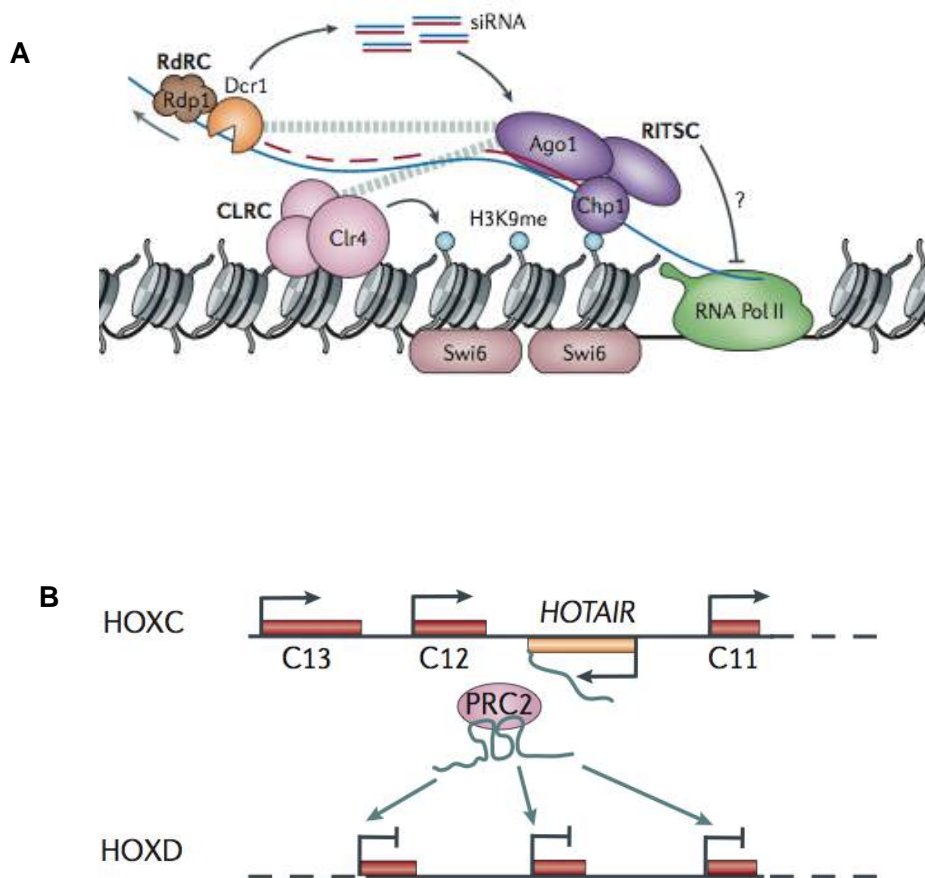


Figure 3. Recrutement des facteurs de l'hétérochromatine et des PcG par des ARN non codants. **A.** Formation de l'hétérochromatine péricentromérique médiée par la voie de l'ARN interférence chez *S. pombe*. La transcription des régions péricentromériques conduit à la formation de duplexes d'ARN qui sont clivés par Dcr1 en siARN. Les siARN sont pris en charge par Ago1 dans le complexe RITS qui va permettre le recrutement de la lysine méthyltransférase Clr4 au niveau des séquences homologues. Clr4 catalyse la méthylation de H3K9 qui sert d'ancrage à la protéine à chromodomaine Swi6, et permet la formation d'une structure hétérochromatinienne. Une autre protéine à chromodomaine, Chp1, fait partie du complexe RITS et contribue à son ciblage en reconnaissant la méthylation de H3K9. **B.** Exemple de recrutement des PcG par un lncARN. Chez l'Homme, la transcription antisens du locus HOXC produit l'ARN non codant *HOTAIR* qui est capable d'interagir avec le complexe PRC2. *HOTAIR* permet le ciblage de PRC2 en *trans* au niveau du locus HOXD.

Castel and Martienssen, *Nat. Rev.*, 2013

Beisel and Paro, *Nat. Rev. Genet.*, 2011

cibles de ROR α indépendamment de son rôle dans la méthylation des histones. Autre exemple, les lysine méthyltransférases G9a et GLP, connues pour méthyliser H3K9, ciblent également le facteur de transcription p53 (Huang et al., 2010).

Le complexe PRC1 comprend quatre protéines majeures (Figure 2B). Chez *Drosophila melanogaster*, la protéine à chromodomaine Polycomb (Pc) fixe la triméthylation de H3K27. Les protéines Polyhoméptic (Ph), Posterior sex combs (Psc) et RING1 viennent compléter le complexe (Francis et al., 2001; Shao et al., 1999). RING1 possède une activité ubiquitin ligase sur l'histone H2A qui est stimulée par son association avec Psc (Buchwald et al., 2006). Des homologues de chacun des composants du complexe PRC1 ont été identifiés chez les mammifères : cinq homologues de Pc : CBX2, CBX4, CBX6, CBX7 et CBX8, deux homologues de RING1 : RING1, RING2, trois homologues de Ph : PHC1, PHC2 et PHC3 et deux homologues de Psc : BMI1 et MEL18. (Di Croce and Helin, 2013; Schwartz and Pirrotta, 2013).

III. Ciblage des facteurs à leurs séquences cibles

La première étape dans l'établissement des différents types de chromatine répressive consiste à recruter les acteurs protéiques présentés précédemment au niveau des séquences d'ADN à inactiver. Comment ces facteurs identifient-ils leurs séquences cibles dans la masse génomique ? Tout n'est pas encore compris dans ce domaine ; les mécanismes impliqués semblent être multiples et variables en fonction du contexte génétique.

1. Recrutement par des ARN non codant

a. Formation d'hétérochromatine médiée par la machinerie de l'ARN interférence

D'abord identifiées comme des mécanismes de régulation post-transcriptionnels et cytoplasmiques des gènes, les voies de l'ARN interférence sont maintenant reconnues pour participer à la régulation transcriptionnelle de nombreuses régions génomiques et notamment à la formation d'hétérochromatine. Les mécanismes de ciblage de l'hétérochromatine par des petits ARN non codants ont été décrits en premier lieu dans les régions péricentromériques de la levure *S. pombe* (Volpe et al., 2002) et font désormais office de paradigme dans ce domaine. Chez *S.*

pombe, les régions péricentromériques sont composées de répétitions inversées qui sont transcrites à partir des brins sens et antisens. Les ARN double brins issus de cette activité transcriptionnelle sont clivés en siARN d'environ 21 nucléotides par l'activité RNase-III de la protéine Dicer 1 (Dcr1). Les siARN s'associent ensuite à la protéine Argonaute 1 (Ago1) et guident le complexe protéique contenant Ago1 (complexe RITS : RNA-induced initiation of transcriptional gene silencing) vers les séquences d'ADN homologues, peut-être par l'intermédiaire d'une interaction ARN-ARN avec des transcrits naissants. Le complexe RITS permet le recrutement de Clr4 (homologue de Su(var)3-9 chez *S. pombe*) qui catalyse la diméthylation de H3K9 dans les régions péricentromériques. La marque est reconnue et fixée par Swi6 (HP1) conduisant à l'assemblage d'une structure hétérochromatinienne (Figure 3A) (Castel and Martienssen, 2013). Le recrutement artificiel de Clr4 dans une région euchromatique conduit à la formation d'hétérochromatine même en absence de la machinerie de l'ARN interférence indiquant que la fonction des siARN est restreinte au ciblage du complexe RITS (Kagansky et al., 2009).

Des mécanismes similaires de formation de l'hétérochromatine par des petits ARN non codants ont été mis en évidence chez d'autres organismes. Chez la Drosophile, la voie des piARN (petits ARN non codants produits dans la lignée germinale) dégrade les transcrits des transposons et dirige la méthylation de H3K9 pour réduire l'activité transcriptionnelle des transposons et empêcher leur mobilisation (Ge and Zamore, 2013). Chez *A. thaliana*, les siARN de 24 nucléotides dirigent la méthylation de l'ADN et la formation d'hétérochromatine en recrutant des ADN méthyltransférases dans les régions répétées du génome. Ils pourraient également être impliqués dans le ciblage de SUVH4 (homologue de SUVAR3-9) pour méthyler H3K9 dans ces régions (Johnson et al., 2007).

b. Ciblage des complexes Polycomb par des ARN non codants

Chez le cilié *Tetrahymena thermophila*, une voie de l'ARN interférence spécifique du développement, impliquant des scanARN de 28 nucléotides et la protéine de type Argonaute Twi1p, est requise pour l'action de EZL1, une lysine méthyltransférase apparentée à Enhancer of zeste (Liu et al., 2007). En revanche, il n'y a pas d'évidence à ce jour que la machinerie de l'ARN interférence soit requise pour le recrutement des protéines Polycomb chez les eucaryotes supérieurs (Hekimoglu and Ringrose, 2009).

Un certain nombre de longs ARN non codants (lncARN) sont cependant associés aux Polycomb et pourraient intervenir dans leur ciblage ou leur ancrage aux séquences d'ADN cibles. D'abord, l'ARN non codant *Xist*, long de 17 kb et célèbre pour son rôle dans l'inactivation du chromosome X chez les mammifères, peut être immunoprécipité avec des composants du complexe PRC2 et peut s'associer avec une forme recombinante de EZH2 (Zhao et al., 2008). L'interaction de *Xist* avec EZH2 s'effectue au niveau d'une séquence de 1,6 kb nommée RepA et permettrait la fixation du complexe PRC2 sur une large portion du chromosome X. Un autre exemple de recrutement par des ARN non codants provient de la mise en évidence d'une interaction entre l'ARN non codant *ANRIL* et Su(z)12, un composant du complexe PRC2. *ANRIL* est un lncARN qui est transcrit à partir du locus INK4a/ARF codant pour des gènes suppresseurs de tumeur. Le locus est régulé par les complexes PcG et la déplétion de *ANRIL* empêche la fixation de Su(z)12 au locus cible suggérant que le lncARN est responsable du recrutement du complexe PRC2 (Kotake et al., 2011). Par ailleurs, il a été montré que la sous-unité CBX-7 du complexe PRC1 interagissait avec *ANRIL*. Cet ARN non codant, conjointement avec la triméthylation de H3K27, pourrait donc également permettre le recrutement du complexe PRC1 au locus INK4a/ARF (Yap et al., 2010). Enfin, l'ARN non codant *HOTAIR*, issu de la transcription antisens du cluster HOXC chez l'Homme, fixe PRC2 et permet son recrutement en *trans* au niveau du locus HOXD (Figure 3B) (Rinn et al., 2007; Tsai et al., 2010). S'il est surexprimé, *HOTAIR* peut aussi cibler le complexe PRC2 vers d'autres sites génomiques qui sont souvent des gènes associés au développement (Gupta et al., 2010).

2. Recrutement par des séquences d'ADN ou des facteurs de fixation à l'ADN

a. Recrutement des complexes PcG par des éléments d'ADN spécifiques et des facteurs d'interaction avec l'ADN

Chez *Drosophila melanogaster*, des éléments d'ADN spécifiques mesurant quelques centaines de paires de bases, appelés PRE pour Polycomb Response Elements, sont responsables du recrutement des complexes Polycomb. La délétion de ses séquences régulatrices cause des phénotypes homéotiques (Chan et al., 1994; Müller and Kassis, 2006). Les PRE peuvent être très distants des gènes cibles mais sont souvent proches des sites de démarrage de la transcription (Kharchenko et al., 2011). Ils présentent peu d'homologie de séquence entre eux mais sont

fréquemment enrichis en différents facteurs de fixation de l'ADN (ex : facteurs GAGA, Zeste, Dsp1...) qui pourraient être capables de recruter les complexes PcG (Déjardin et al., 2005; Hodgson et al., 2001; Horard et al., 2000), conjointement avec d'autres mécanismes.

Chez les cellules souches embryonnaires de mammifères, le facteur JARID2, connu pour son implication dans le développement embryonnaire précoce, interagit et colocalise avec les composants du complexe PRC2 au niveau de la plupart des gènes cibles des PcG (Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009). JARID 2 interagit avec l'ADN par l'intermédiaire de son domaine ARID (AT-rich interactive domaine) qui est également requis pour le recrutement du complexe PRC2. Les effets de la déplétion de JARID2 sur l'activité de ce complexe ne sont pas très clairs. Il semble l'absence que l'absence de JARID2 n'ait qu'un effet léger sur le niveau de méthylation de H3K27 suggérant que ce facteur participe au recrutement des complexes PcG en coordination avec d'autres mécanismes. (Li et al., 2010; Pasini et al., 2010).

b. Formation d'hétérochromatine par des facteurs spécifiques de fixation à l'ADN

Comme évoqué précédemment, chez la levure *S. pombe*, des régions d'hétérochromatine constitutive se forment dans les régions péricentromériques mais également au niveau des télomères et au locus du mating type, grâce à la machinerie de l'ARN interférence (Buhler and Gasser, 2009; Grewal, 2010). En plus de cette voie, des facteurs reconnaissant des séquences d'ADN spécifiques peuvent recruter des protéines de l'hétérochromatine. Au locus du mating type, les facteurs de transcription Atf1 et Pcr1, ainsi que les protéines de fixation aux télomères Taz1 et Ccq1 permettent d'initier la formation d'hétérochromatine par le recrutement de l'histone désacétylase Clr3. Celle-ci attire la lysine méthyltransférase Clr4 (homologue de SU(VAR)3-9), et Swi6 (homologue de HP1), conduisant à la nucléation et à la diffusion de structures hétérochromatiques (Jia et al., 2004; Kanoh et al., 2005; Yamada et al., 2005).

La déposition de HP1 médiée par des facteurs de transcription permet aussi d'établir des environnements d'hétérochromatine nécessaires à la répression de la transcription génique. HP1 est notamment impliquée dans la régulation transcriptionnelle des protéines KRAB-ZPF (Krüppel-associated box domain zinc-finger proteins) qui constituent un important groupe de régulateurs transcriptionnels dans les génomes des eucaryotes supérieurs (Sripathy et al., 2006).

Cette classe de gènes est transcriptionnellement auto-régulée par la fixation de protéines KRAB-ZPF spécifiques à l'extrémité 3' des gènes KRAB-ZPF qui doivent être réprimés (O'Geen et al., 2007). Celles-ci recrutent le facteur KAP1 (KRAB-associated protein 1) qui sert d'ancrage pour différents facteurs de l'hétérochromatine comme HP1 et SETDB1, une lysine méthyltransférase spécifique de H3K9. Le recrutement de KAP1 conduit à la diffusion de l'hétérochromatine grâce à la propagation de H3K9me3 et à la fixation de HP1 sur tout le corps des gènes ciblés, permettant ainsi la répression des promoteurs qui peuvent se trouver à plusieurs dizaines de kb en amont du site de nucléation (Groner et al., 2010; Vogel et al., 2006)

IV. Assemblage des structures chromatinienne répressives et inhibition de l'activité transcriptionnelle des séquences cibles

Une fois recrutées à leurs séquences cibles, les complexes PRC2 et SU(VAR)3-9 catalysent respectivement la méthylation de H3K27 et de H3K9 par l'intermédiaire de leur activité lysine méthyltransférase. La fixation de ces marques d'histones par les protéines à chromodomaine, Pc (pour H3K27me2/me3) ou HP1 (pour H3K9me2/me3), est généralement associée à la condensation de la chromatine et à la réorganisation des domaines nucléaires. Cela conduit à la réduction de l'accessibilité de la machinerie transcriptionnelle et d'autres complexes enzymatiques.

1. Condensation de l'hétérochromatine

La condensation de la chromatine par les protéines HP1 semble s'établir essentiellement grâce à la dimérisation des molécules HP1 fixées sur des nucléosomes adjacents, par l'intermédiaire de leur domaine chromoshadow (Brasher et al., 2000; Verschure et al., 2005) (Figure 4A). Cependant, les protéines HP1 peuvent également servir de plate-forme pour d'autres enzymes de modification de la chromatine qui peuvent, elles aussi, empêcher la machinerie transcriptionnelle d'accéder aux séquences sous-jacentes. Chez *S. pombe*, la protéine Clr3 se répand sur l'hétérochromatine grâce à son activité histone désacétylase (HDAC) et grâce à Swi6. Clr3 favorise la maintenance de la structure chromatinienne en stabilisant la triméthylation de H3K9 et en limitant l'accès de l'ARN polymérase II (Yamada et al., 2005). Chez les eucaryotes multicellulaires, des mécanismes similaires pourraient être impliqués et pourraient certainement agir en coopération avec la

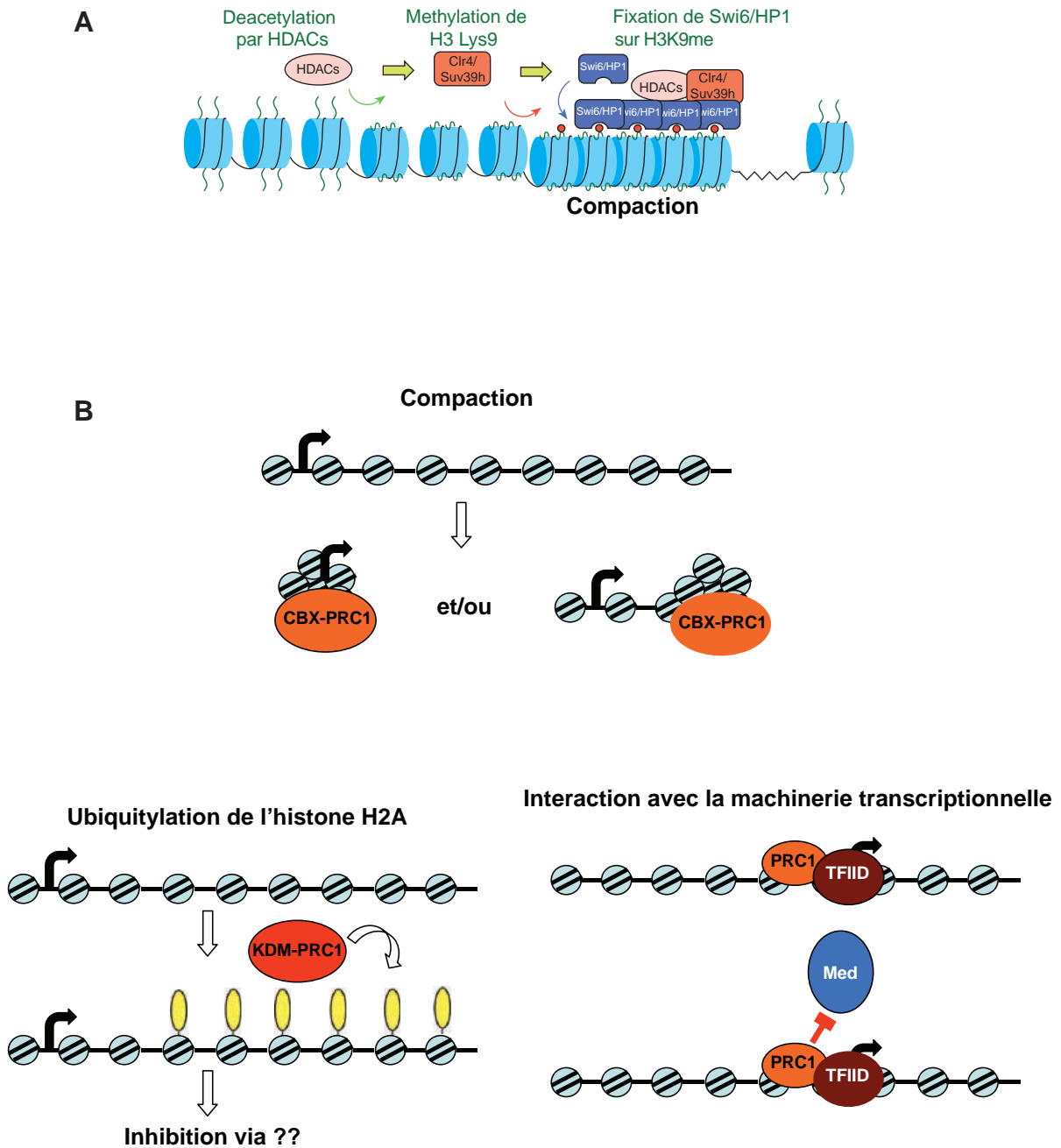


Figure 4. Mécanismes de répression par HP1 et par le complexe PRC1. A. La fixation de HP1 sur les nucléosomes méthylés par les protéines SU(VAR)3-9 entraîne la compaction des nucléosomes de la région ciblée. L'action d'histone deacétylases renforce la structure hétérochromatique. B. La fixation du complexe PRC1 entraîne l'inhibition de la transcription dans les régions ciblées 1. par la compaction des nucléosomes, 2. par l'ubiquitylation de l'histone H2A, 3. par l'interaction directe avec la machinerie de la transcription (blocage du complexe Mediator requis pour l'activation de la transcription)

Adapté de Elgin and Grewal, *Curr Biol.*, 2003 et de Simon and Kingston, *Mol Cell Rev.*, 2013

méthylation de l'ADN, une marque également impliquée dans la régulation transcriptionnelle. Chez les mammifères, la méthylation de l'ADN recrute des protéines de fixation aux methyl-CpG (MBD proteins) qui, de manière analogue à HP1, servent de plate-forme à des enzymes de modification de la chromatine comme les HDAC (Bird and Wolffe, 1999).

2. Répression transcriptionnelle par les protéines PcG

Le complexe PRC1 conduit également à la condensation de la chromatine (Figure 4B). La compaction, *in vitro* par le complexe PRC1 d'un filament de chromatine en structures plus denses est visualisable en microscopie électronique (Francis et al., 2004; Grau et al., 2011). La chromatine associée aux complexes PcG est résistante à l'activité de remodelage de la chromatine par les protéines de type SWI/SNF (Francis et al., 2001). Les propriétés de compaction et d'inhibition du remodelage sont des fonctions non enzymatiques du complexe PRC1 qui sont associées au composant Psc chez la Drosophile et chez *C. elegans* ou à la sous-unité Pc chez les vertébrés. De manière générale, la capacité de condensation des nucléosomes semble attribuable à des domaines très enrichis en acides aminés de charge positive (Grau et al., 2011). De manière intéressante, il a été montré que l'état condensé de la chromatine stimulait la méthylation de H3K27 par le complexe PRC2 (Yuan et al., 2012). Comme le chromodomaine du complexe PRC1 fixe H3K27me3, il pourrait exister une boucle de régulation positive entre la compaction des nucléosomes et la marque d'histone qui renforcerait la stabilité de la structure chromatinienne.

L'ubiquitylation de H2A (H2AK119ub1) par les sous-unités RING1 et Psc du complexe PRC1 joue également un rôle important dans le mécanisme de répression par les Polycomb. Cette modification d'histone intervient dans l'inhibition transcriptionnelle de certains gènes mais est requise à différents degrés en fonction des cibles. Par exemple, dans des cellules souches embryonnaires de souris, la présence d'une protéine RING1B incapable de catalyser l'ubiquitylation conduit à la dérégulation presque complète des gènes *ZIC1* et *PAX3* tandis que de nombreux gènes *HOX* sont encore maintenus réprimés. L'activité d'ubiquitylation du complexe PRC1 n'est pas requise pour sa fixation aux gènes cibles ou pour son activité de compaction de la chromatine (Endoh et al., 2012). A ce jour, les mécanismes par lesquels H2AK119ub1 conduit à la répression sont encore inconnus. (Figure 4B)

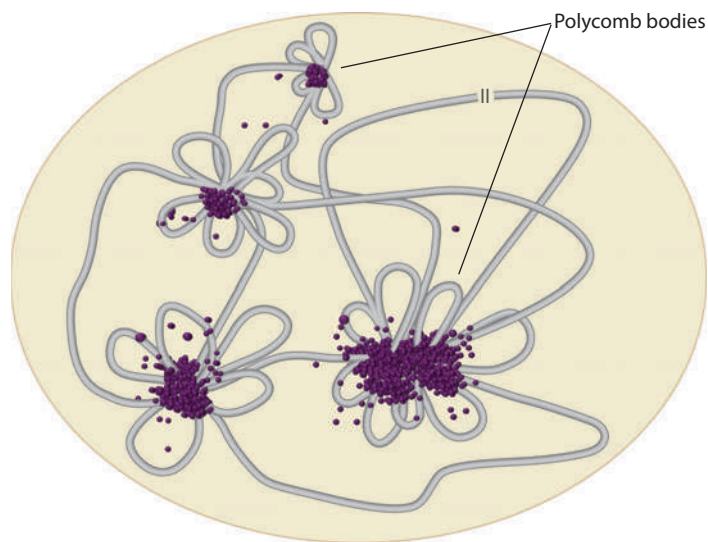


Figure 5. Représentation schématique des «Polycomb bodies» dans le noyau. Les protéines PcG sont concentrées dans certaines zones du noyau. Plusieurs gènes cibles, parfois très distants sur l'ADN génomique sont regroupés dans un même PcG body. Ces structures nucléaires pourraient renforcer l'efficacité de la répression médiée par les PcG.
Lanzuolo and Orlando., Annu Rev Genet., 2012

Enfin, les complexes PcG pourraient réprimer leurs séquences cibles en interagissant directement avec la machinerie transcriptionnelle et en inhibant sa fonction. Il a notamment été observé que le complexe PRC1 pouvait empêcher la transcription d'une matrice d'ADN nue *in vitro* (King et al., 2002). De plus, il a été observé dans des cellules ES de souris que les gènes ciblés par PRC1 et PRC2 présentaient moins de promoteurs associés à l'ARN polymérase II que la moyenne des gènes (Min et al., 2011), ce qui serait cohérent avec un blocage de l'initiation de la transcription par les complexes PRC à ces promoteurs. Une étude récente a permis de mieux comprendre les mécanismes d'action des protéines PcG en montrant que le complexe PRC1 inhibait la fixation du complexe Mediator, un composant clé de l'activation de la transcription (Lehmann et al., 2012) (Figure 4B).

V. Formation de structures nucléaires d'ordre supérieur

Dans le noyau, l'organisation des domaines chromatinien n'est pas aléatoire. Selon les modèles actuels, la position des séquences d'ADN dans le volume nucléaire pourrait influencer leur expression en régulant leur accès à différents facteurs comme les machineries de transcription et de réplication (Mao et al., 2011). L'organisation des domaines chromatinien répressifs en structures nucléaires d'ordre supérieur pourrait influencer la position des séquences dans le noyau et contribuer à leur répression. L'organisation des domaines nucléaires pourrait aussi avoir une fonction structurale pour le maintien de l'intégrité des chromosomes.

1. Cluster hétérochromatiques

Chez *S. pombe*, les télomères s'assemblent pour former 2 à 4 foci à la périphérie nucléaire. La protéine télomérique Taz1, impliquée dans la formation de l'hétérochromatine ainsi que la machinerie de l'ARN interférence sont requises pour le clustering des télomères (Cooper et al., 1998; Hall et al., 2003). Chez la Drosophile, des chromosomes polytènes contenant des sites de fixation de HP1 ectopiques peuvent former des boucles qui associent ces sites à l'hétérochromatine péricentromérique et télomérique. HP1, avec d'autres facteurs, pourrait donc faciliter la coalescence de régions hétérochromatiques dispersées et organiser des structures nucléaires de répression chromatinienne (Seum et al., 2001).

2. Polycomb bodies

Des observations cytologiques ont montré, chez la *Drosophile* et chez l'Homme que les protéines des complexes PcG étaient concentrées en foci. Ces domaines nucléaires distincts, appelés « Polycomb bodies » sont souvent localisés dans des zones proches de l'hétérochromatine péricentromérique (Cmarko et al., 2003) (Figure 5). Les gènes Hox sont fréquemment retrouvés dans les Polycomb bodies quand ils sont réprimés, suggérant que cette organisation nucléaire est impliquée dans le mécanisme de répression par les PcG (Bantignies et al., 2011). Le nombre de foci PcG est cependant moins important que le nombre de gènes cibles des Polycomb indiquant que plusieurs de ces gènes doivent être regroupés dans les mêmes PcG bodies. En effet, des expériences de FISH menées chez la *Drosophile* ont montré des interactions entre différents PRE du même cluster de gènes HOX et entre des loci HOX distant de plusieurs mégabases (Bantignies et al., 2011; Cléard et al., 2006; Lanzuolo et al., 2007). Des analyses par 3C (Chromosome Conformation Capture) ont également apporté l'évidence d'interactions de longue distance entre les gènes cibles de PcG (Lanzuolo et al., 2007).

En conclusion, la mise en place de structures chromatinienne répressives repose sur la coexistence de mécanismes de compaction de la chromatine et d'interactions à longue distance de ces domaines chromatinien. Les voies qui permettent le recrutement des facteurs nécessaires à l'établissement des domaines chromatinien répressifs sont multiples. Si le ciblage des facteurs de l'hétérochromatine dépend essentiellement de la machinerie de l'ARN interférence, les mécanismes de recrutement des complexes Polycomb semblent plus hétérogènes et multifactoriels.

CHAPITRE 2

LES CENTROMERES, ACTEURS DU MAINTIEN DE L'INTEGRITE DES GENOMES

Intéressons-nous maintenant à une autre région chromatinienne particulièrement importante pour le maintien de la stabilité des génomes eucaryotes, le centromère. Dans le noyau des organismes eucaryotes, l'ADN génomique est organisé en différents chromosomes physiquement distincts les uns des autres. Les centromères sont des régions chromosomiques cruciales pour le maintien de l'intégrité de ces génomes puisqu'ils permettent la ségrégation fidèle des chromosomes à chaque division cellulaire et assurent le maintien des fonctions germinales, zygotiques et somatiques. Des défauts de l'activité des centromères peuvent générer l'arrêt du cycle cellulaire, de la mort cellulaire ou de l'aneuploïdie (présence d'un nombre de copies incorrect de certains chromosomes) qui peut se traduire par de l'infertilité ou par la transformation tumorale des cellules concernées (Weaver and Cleveland, 2007). Les régions centromériques constituent une plate-forme pour l'assemblage d'une structure protéique particulière, le kinétochore, qui interagit avec les microtubules du fuseau mitotique et permet le mouvement des chromosomes à chaque pôle du fuseau (Cheeseman and Desai, 2008). En dépit de leur fonction essentielle et conservée chez les eucaryotes, les centromères présentent une importante diversité de structure et montrent une étonnante plasticité. En effet, en réponse à des réarrangements chromosomiques, les centromères peuvent être inactivés, réactivés et même, assemblés dans des régions ectopiques (Allshire and Karpen, 2008). Cette fluidité est permise grâce à une régulation épigénétique de l'activité des centromères dont tous les mystères n'ont pas encore été complètement percés.

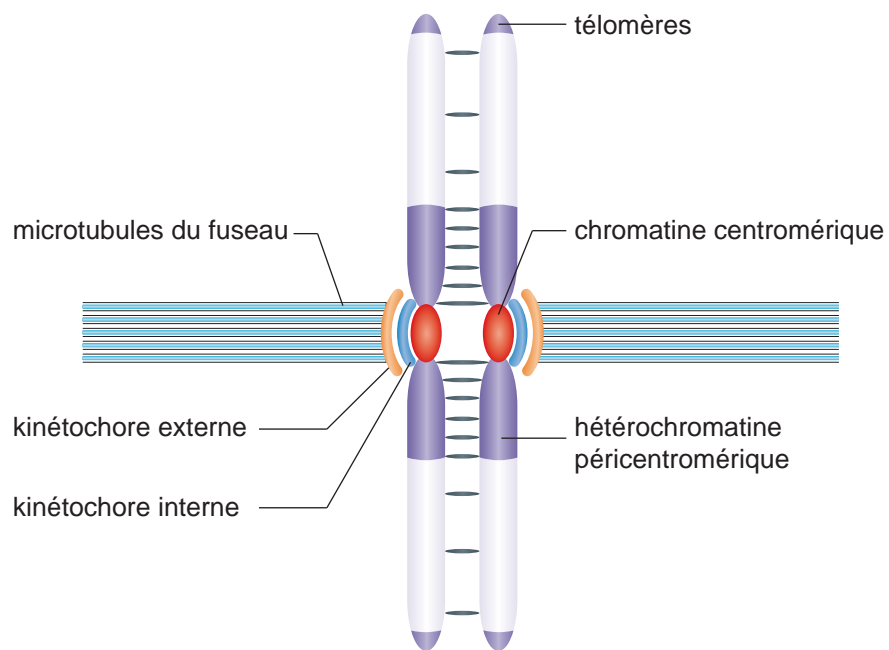


Figure 6. Structure générale des centromères (chromosome monocentrique). Au niveau des centromères, les kinétochores internes et externes s'assemblent sur la chromatine centromérique. Le kinétochore externe interagit avec les microtubules du fuseau mitotique. L'hétérochromatine péricentromérique est riche en cohésine ce qui assure la cohésion des chromatides sœurs.

Adapté de Allshire and Karpen, Nat. Rev., 2008

I. Le centromère, structure générale

Les centromères sont constitués d'une région d'ADN centromérique sur laquelle s'assemble un complexe protéique, le kinétochore, qui a pour but d'interagir avec les microtubules du fuseau mitotique pour permettre la ségrégation fidèle des chromosomes à chaque division cellulaire. Chez la plupart des systèmes eucaryotes modèles, les protéines du kinétochore s'assemblent en une seule région des chromosomes. Sur ces chromosomes dits « monocentriques », le centromère est visible en métaphase sous forme d'une constriction qui peut être localisée sur n'importe quelle zone du chromosome. Chez certains organismes cependant, les protéines du kinétochore peuvent se fixer sur la toute la longueur des chromosomes. Ces chromosomes sont qualifiés de « holocentriques » et sont retrouvés dans une grande variété d'organismes y compris les plantes, différents animaux comme les nématodes, les arachnides et plusieurs groupes d'insectes (Melters et al., 2012). En dépit d'une structure différente, les protéines du kinétochore sont conservées entre les organismes à chromosomes monocentriques et les organismes à chromosomes holocentriques (Dernburg, 2001).

La structure du kinétochore est historiquement divisée en deux régions distinctes : le kinétochore interne, qui se forme à l'interface de l'ADN centromérique et le kinétochore externe, une région de 50 à 60 nm d'épaisseur qui établit une surface d'interaction pour les microtubules (Figure 6). Actuellement, environ 80 protéines du kinétochore ont été identifiées chez l'Homme. Bien qu'il y ait des différences entre les organismes, les éléments clés de la composition et de l'organisation du kinétochore sont conservés parmi les eucaryotes. Dans le kinétochore interne, un certain nombre de protéines sont associées constitutivement avec l'ADN centromérique et constituent le « Constitutive Centromere-Associated Network » (CCAN). Le CCAN est composé d'un variant d'histone H3 centromérique, nommé Cenp-A chez l'Homme, qui s'intègre dans l'ADN centromérique, de la protéine centromérique associée à l'ADN Cenp-C et, chez les vertébrés, de 13 autres facteurs d'interaction (Cenp-H, Cenp-I, Cenp-K – U) (Cheeseman, Desai, 2008). Cenp-A et Cenp-C sont des constituants essentiels du kinétochore qui sont retrouvés chez tous les eucaryotes étudiés à ce jour (Heeger et al., 2005; Oegema et al., 2001; Saitoh et al., 1992). Les autres constituants du CCAN n'ont pas été identifiés chez *D. melanogaster*, *C. elegans* ou chez les plantes. Les facteurs constitutifs permettent le recrutement de protéines additionnelles au kinétochore pendant la fin de la phase G2 et pendant la prophase de mitose. Parmi ces facteurs du kinétochore externe, on retrouve notamment des composants du point de contrôle mitotique

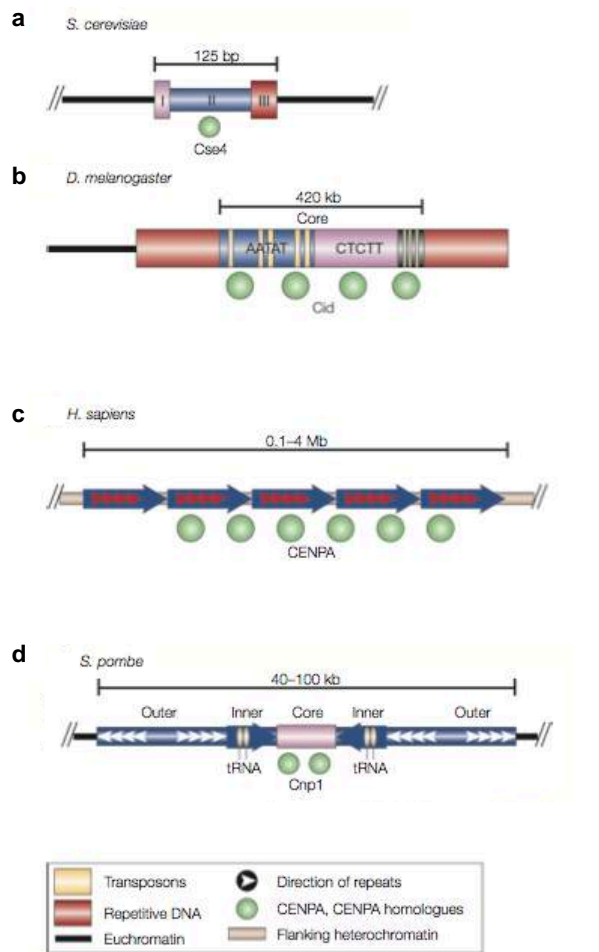


Figure 7. ADN centromériques de différents eucaryotes. La séquence des centromères n'est pas conservée entre les espèces. **a.** La fonction des centromères de *Saccharomyces cerevisiae* dépend d'une séquence de 125 pb qui contient trois éléments uniques (I, II, III). **b.** Chez *Drosophila melanogaster*, les centromères sont composés de séquences répétées et de transposons. **c.** Les centromères humains (*Homo sapiens*) sont constitués d'ADN alpha-satellite contenant des répétitions en tandem qui peuvent s'étendre sur plusieurs Mb. **d.** les centromères de *Schizosaccharomyces pombe* contiennent une région centrale unique, flanquée par des répétitions inversées internes et externes. (a-d) Tous les centromères sont associés au variant d'histone CenH3 (homologues de Cenp-A).

Adapté de Sullivan et al., Nat Rev Genet., 2001

(Mad1, Mad2, BubR1), le complexe d'attachement aux microtubules Ndc80 et la protéine à domaine coiled-coil Cenp-F. Ces facteurs seront ensuite délocalisés dès la sortie de mitose des cellules (Hori et al., 2003; Liu et al., 2003, 2006).

Les kinétochores doivent s'établir dans des régions chromosomiques bien spécifiées pour assurer leur fonction. Comprendre les mécanismes qui régulent l'activité des centromères demande de comprendre quels sont les déterminants de l'identité des centromères. Qu'est-ce qui fait d'une région chromosomique un centromère ?

II. Déterminants de l'identité centromérique

1. L'ADN centromérique

La nature essentielle des centromères et leur localisation stable au sein des cellules d'un même organisme pourraient suggérer, à première vue, que les centromères sont des régions chromosomiques définies par une séquence d'ADN particulière. C'est effectivement le cas chez la levure *S. cerevisiae* chez qui la fonction centromérique est spécifiée par une région d'ADN de seulement 125 pb qui est composée de trois éléments uniques : CDEI, CDEII et CDEIII (CDEs : Centromere DNA Elements). L'élément CDEIII de 15 pb permet l'attraction d'un complexe contenant des protéines de fixation de l'ADN séquence-spécifique (Ndc10, Cep3, Ctf13 et Skp1). Ce complexe dicte l'assemblage d'un nucléosome unique, contenant le variant d'histone centromérique Cse4 (homologue de Cenp-A), qui permet le recrutement des autres protéines du kinétochore (Westermann et al., 2007). La mutation d'une seule base dans le domaine CDEII abolit l'incorporation de Cse4 et conduit à la perte de la fonction centromérique et à la mort des cellules (Cottarel et al., 1989; Spencer and Hieter, 1992).

Cette dépendance à une séquence d'ADN pour établir une fonction centromérique ne se retrouve cependant pas chez les autres eucaryotes. D'une part, il a été observé dans de nombreux systèmes que l'ADN centromérique n'était ni nécessaire ni suffisant à l'établissement d'un centromère actif. Chez l'Homme et la *Drosophile* par exemple, des kinétochores actifs peuvent se former dans des régions ectopiques qui ne contiennent pas de séquences homologues à celles des centromères normaux. Une fois établis, ces « néocentromères » maintiennent parfaitement leur fonction centromérique au cours des divisions cellulaires (Warburton, 2004; Williams et al., 1998). A

l'inverse, la présence d'ADN centromérique ne garantit pas l'assemblage d'un kinétochore fonctionnel puisque des centromères peuvent être inactivés sans altération de leur séquence d'ADN (voir IV) (Agudo et al., 2000; Earnshaw and Migeon, 1985). D'autre part, il est étonnant de noter que même si la fonction des centromères est très conservée entre les organismes, la taille et la nature de l'ADN centromérique peuvent être très variables (Figure 7). Chez la levure *S. pombe*, les centromères sont composés d'une région centrale constituée de séquences uniques de 4 à 5 kb auxquelles s'associe le variant d'histone Cnp1 (homologue de Cenp-A). Cette région est flanquée par des répétitions inversées internes et externes, partagées par les trois chromosomes de la cellule. L'ensemble de la région centromérique constitue 40 à 100 kb (Clarke et al., 1986; Nakaseko et al., 1986). Les centromères de *Drosophila melanogaster* sont quant à eux composés de répétitions de transposons et de séquences satellites (AATAT et CTCTT) qui peuvent mesurer jusqu'à 420 kb (Sun et al., 2003). Chez la levure *Candida albicans* en revanche, les centromères sont des séquences uniques et différentes sur chaque chromosome (Sanyal et al., 2004). Chez *Arabidopsis thaliana*, les centromères sont contenus dans des régions de 500 à 1000 kb qui contiennent des répétitions de 180 bp et qui sont bordées par des séquences d'ADN ribosomiques ou d'autres séquences répétées (Murata et al., 1994; Round et al., 1997). Les centromères de différentes espèces de betterave (*genus Beta*) ont également été décrits et comprennent des régions d'ADN satellite et d'éléments transposables (Gindullis et al., 2001). Enfin, chez l'Homme, les centromères sont constitués d'ADN alpha-satellite riche en nucléotides A et T et composé de monomères de 171 pb répétés en tandem. Ces derniers s'assemblent en répétitions d'ordre supérieur qui s'étendent sur quelques centaines de kb à plusieurs Mb (Mitchell et al., 1985; Willard, 1985). La Figure 7 présente quelques exemples de séquences d'ADN centromérique.

Si l'ADN centromérique ne semble pas être un déterminant clé de l'identité des centromères chez la plupart des espèces, la présence récurrente de séquences répétées et d'ADN satellite dans les régions centromériques suggère que la composition des séquences centromériques pourrait tout de même favoriser l'installation des kinétochores. En outre, chez l'Homme, des séquences de 17 pb appelées Cenp-B box, reconnues par la protéine centromérique Cenp-B, sont, avec l'ADN alpha-satellite, suffisantes à la formation *de novo* des centromères (Ohzeki et al., 2002). A l'heure actuelle, le rôle de l'ADN dans la détermination de l'identité centromérique n'est pas complètement clarifié.

2. Cenp-A/CenH3 : marqueur des centromères actifs

L'identité centromérique est en réalité définie épigénétiquement par une composition chromatinienne particulière. Chez les eucaryotes, tous les centromères actifs, capables de mobiliser un kinétochore fonctionnel, sont assemblés au niveau de régions chromatiniennes dans lesquelles l'histone canonique H3 a été remplacée par le variant d'histone H3 centromérique, composant universel du CCAN : CenH3. D'abord identifié chez l'Homme et nommé Cenp-A, ce variant d'histone a ensuite été identifié chez de nombreux autres modèles eucaryotes : Cse4 chez la levure *S. cerevisiae*, Cnp1 chez *S. Pombe*, CID chez *Drosophila melanogaster* ou encore HCP-3 chez *C. elegans* (Blower and Karpen, 2001; Buchwitz et al., 1999; Earnshaw and Rothfield, 1985; Stoler et al., 1995; Takahashi et al., 2000). Plus généralement, le variant d'histone centromérique est nommé CenH3 pour « Centromeric Histone H3 ». Les protéines CenH3 sont retrouvées au niveau de tous les centromères fonctionnels, notamment au niveau des néocentromères mais sont absentes des centromères inactivés (Earnshaw and Migeon, 1985; Saffery et al., 2000). Elles sont des protéines fondatrices pour l'assemblage des autres composants du kinétochore. En effet, la déplétion des protéines CenH3 conduit à la délocalisation de la plupart des protéines du kinétochore alors que la suppression des autres protéines du kinétochore n'a pas d'effet sur la localisation de CenH3 (Blower and Karpen, 2001; Collins et al., 2005; Howman et al., 2000; Liu et al., 2006; Oegema et al., 2001). De plus, chez la Drosophile, la surexpression de CenH3 conduit à sa délocalisation dans des régions non-centromériques et à la formation de kinétochores ectopiques (Heun et al., 2006). Le variant d'histone CenH3 est donc l'un des facteurs fondateurs du kinétochore et un déterminant essentiel de la fonction des centromères.

Les caractéristiques qui procurent à CenH3 ses propriétés centromériques ne sont pas complètement comprises. Cependant, d'importantes différences de séquence et de structure existent entre H3 et son variant centromérique CenH3. Comme toutes les protéines histones, CenH3 contient un domaine globulaire appelé histone fold domain (HFD) qui est composé de quatre hélices alpha (α N, α 1, α 2 et α 3) séparées par des boucles (L1 et L2). Ce domaine HFD est assez divergent de celui des histones H3 et contient quelques acides aminés supplémentaires dans la boucle L1. Chez l'Homme par exemple, Cenp-A ne présente que 60% de similarité avec H3 dans le domaine HFD et contient une insertion de deux acides aminés dans la boucle L1 (Sullivan et al., 1994b) (Figure 8A). Le domaine HFD de CenH3 procure aux nucléosomes des

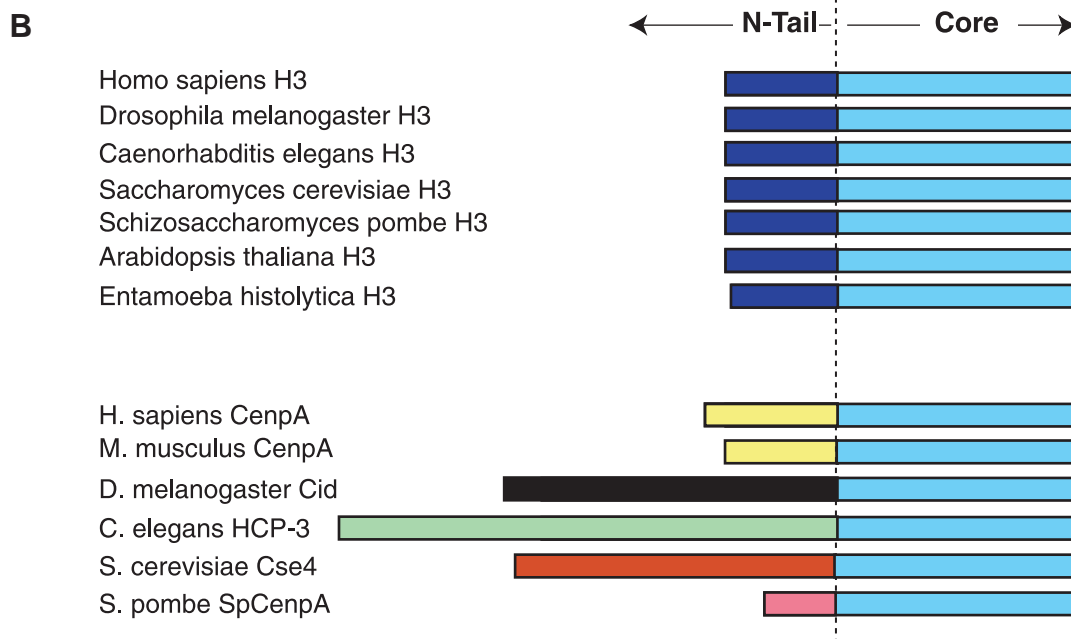
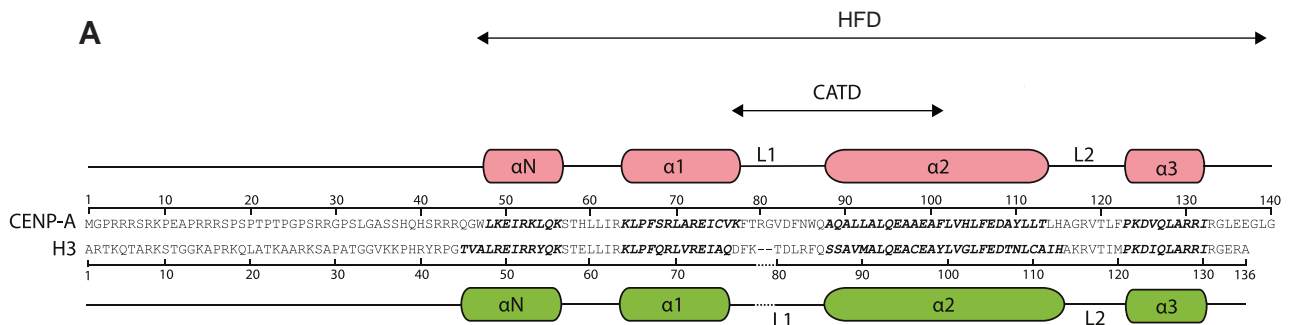


Figure 8. Structure des protéines CenH3. **A.** Séquences protéiques primaires de l'histone centromérique humaine Cenp-A et de l'histone canonique H3.1. Le domaine globulaire (HFD) est composé de quatre hélices alpha séparées par les boucles L1 et L2. La boucle 1 de Cenp-A contient deux acides aminés supplémentaires par rapport à celle de H3. La région N-terminale de Cenp-A est très divergente de celle de H3. **B.** Alignement schématique des protéines CenH3 de différents organismes. Les domaines globulaires des histones centromériques sont similaires mais les régions N-terminales sont très divergentes en séquence et en longueur entre les différentes espèces.

Adapté de Stellfox et al., Cell Mol Life Sci.2013 et de Henikoff et al., Science, 2001

propriétés physiques distinctes de celles des nucléosomes canoniques qui pourraient contribuer à des différences fonctionnelles. Un certain nombre d'analyses *in vitro*, réalisées sur des tétramères d'histones ou sur des nucléosomes, ont montré que la structure de CenH3 formait une interface plus rigide et plus compacte avec H4 que l'histone canonique H3. La région responsable de cette différence est localisée dans la première boucle (L1) et dans la seconde hélice alpha ($\alpha 2$) et correspond à des résidus qui sont aussi impliqués dans le recrutement de CenH3 au niveau des centromères (présentés plus bas) (Black et al., 2004).

Les différences entre H3 et CenH3 sont encore plus importantes dans la région N-terminale. Chez l'Homme, Cenp-A ne partage que 24% d'identité avec H3 dans la queue N-terminale (Sullivan et al., 1994b). Malgré la fonction conservée des protéines CenH3, cette région est très divergente en séquence et en longueur entre les organismes. L'extrémité N-terminale contient par exemple 130 acides aminés chez la Drosophile, 45 acides aminés chez l'Homme et 20 acides aminés chez *S. pombe* (Figure 8B) (Henikoff and Dalal, 2005). Peu de choses sont connues sur le rôle de cette région, d'autant plus qu'elle ne contient pas les lysines bien caractérisées de l'histone H3 sujettes à des modifications post-traductionnelles. Une étude récente, menée dans des cellules humaines montre cependant que la queue N-terminale de Cenp-A est soumise à des modifications et notamment à la phosphorylation des sérines 16 et 18. Ces résidus modifient la structure secondaire de la région N-terminale ce qui favorise les interactions intra et intermoléculaires et influence la conformation des nucléosomes CenH3 (Bailey et al., 2013).

Les protéines CenH3 contribuent donc à l'établissement d'une structure chromatinienne particulière qui sert de fondation à la construction des kinétochores. Cependant, une question reste entière : si CenH3 n'est pas recrutée aux centromères par une séquence d'ADN particulière, alors comment est régulée, épigénétiquement, la fonction centromérique ? Comment CenH3 est-elle ciblée et maintenue en une région unique des chromosomes ? Les facteurs impliqués dans la régulation épigénétique de la fonction centromérique sont nombreux et variés et l'ensemble des mécanismes mis en œuvre sont loin d'être élucidés.

III. Régulation épigénétique de l'activité des centromères, un secret bien gardé...

1. Le rôle de CenH3 dans son propre recrutement aux centromères

Les caractéristiques structurales de CenH3 pourraient, en plus de participer à la formation d'une structure chromatinienne particulière, contribuer au ciblage de la protéine aux centromères. Le domaine HFD joue en effet un rôle important dans la localisation du variant centromérique. Une protéine chimérique contenant 22 acides aminés de la boucle 1 (L1) et de l'hélice $\alpha 2$ de la protéine humaine Cenp-A se localise aux centromères (Black et al., 2004). Le ciblage aux centromères de la protéine de fusion se produit même après la déplétion de la protéine CenH3 endogène par ARN interférence. Cela suggère que le recrutement de nouvelles protéines Cenp-A aux centromères est indépendant de la pré-existence de CenpA. Des études similaires menées chez la *Drosophile* ont montré que le domaine requis pour le ciblage de CID aux centromères est encore plus court et contient la partie C-terminale de l'hélice $\alpha 1$, la boucle 1 et la région N-terminale de l'hélice $\alpha 2$ (15 acides aminés) (Vermaak et al., 2002). Cette région, requise pour le recrutement de CenH3 au niveau de l'ADN centromérique a été appelée CATD (Cenp-A Targeting Domain). Des substitutions d'un seul acide aminé dans le CATD empêchent la localisation aux centromères. Comme mentionné plus haut, le CATD interagit de manière particulière avec H4 mais il est peu probable que cette interaction soit suffisante pour le ciblage de CenH3 puisque H4 est présent sur la totalité du génome. D'autres facteurs doivent donc être impliqués dans le ciblage de CenH3.

La composition des nucléosomes contenant CenH3 serait également un facteur important pour le ciblage et le maintien de CenH3 aux centromères au cours de divisions cellulaires. Des expériences de purification de nucléosomes chez l'Homme et la *Drosophile* ont révélé que la majorité des nucléosomes contenant le variant CenH3 devaient exister sous la forme d'octamères [Cenp-A-H4-H2A-H2B]₂ (Blower et al., 2002; Foltz et al., 2006). Cependant, certaines études ont identifié des nucléosomes de composition particulière. Chez *S. cerevisiae*, Cse4 pourrait être assemblé dans un hexamère contenant [Cse4-H4]₂ et deux copies de la protéine non-histone Scm3 en remplacement de H2A et H2B (Mizuguchi et al., 2007). Chez la *Drosophile*, CID pourrait être présent dans des tétramères ou hémisomes [CID-H4-H2A-H2B] (Dalal et al., 2007). La composition du nucléosome CenH3 pourrait être dynamique au cours du cycle cellulaire et les nucléosomes commuteraient entre octamères et tétramères (Bui et al., 2012;

Shivaraju et al., 2012). Cette caractéristique pourrait permettre la conservation de l'identité des centromères au cours des divisions cellulaires. En effet, les protéines CenH3 sont généralement chargées en dehors de la phase S de réplication, posant un problème pour l'identité des centromères après le passage de la fourche de réplication (Jansen et al., 2007; Schuh et al., 2007). Il a donc été proposé un modèle dans lequel le nucléosome Cenp-A serait partagé en deux sur chaque brin fils après la réplication et resterait sous forme de tétramère jusqu'à l'incorporation de nouvelles protéines CenH3. L'histone chaperone HJURP/Smc3 (présentée plus bas) pourrait stabiliser les tétramères. Ceci permettrait de conserver la taille de la région centromérique et de préserver la marque épigénétique des centromères au cours des divisions. Un autre modèle postule quant à lui que les nucléosomes CenH3 seraient maintenus sous forme d'octamères et répartis aléatoirement entre les deux brins fils après le passage de la machinerie de réplication. Il a notamment été montré que le variant d'histone H3.3, par ailleurs impliqué dans la transcription, pouvait s'incorporer dans les espaces libres au moment de la phase S avant d'être remplacé par de nouvelles protéines CenH3 (Dunleavy et al., 2011) (Figure 9).

2. Rôle des facteurs d'interaction avec CenH3

Comment restreindre l'incorporation de CenH3 au niveau des séquences centromérique ? Il y a quelques années, le complexe Mis18 constitué de la protéine KNL-2 et de ses partenaires Mis18 α/β a été identifié comme étant requis pour la localisation de CenH3 aux centromères chez un certain nombre d'organismes modèles (Fujita et al., 2007; Hayashi et al., 2004; Maddox et al., 2007). KNL2 contient un domaine SANT, communément retrouvé dans les protéines de fixation de l'ADN et les complexes de remodelage de la chromatine (Maddox et al., 2007). Le recrutement de ce complexe aux centromères à la fin de l'anaphase est l'événement connu le plus en amont de la déposition de CenH3. Il permettrait le « licensing » des centromères. Le « licensing » fait partie de ces mots difficiles à traduire en français... mais d'une certaine façon, le complexe Mis18 transmettrait une sorte d'« autorisation » au niveau des centromères pour l'incorporation de CenH3 (Fujita et al., 2007; Maddox et al., 2007). Cependant, il n'est toujours pas compris comment les facteurs de licensing reconnaissent et fixent la chromatine centromérique avec une forte spécificité pour recruter les composants nécessaires au chargement de CenH3 dans l'ADN centromérique.

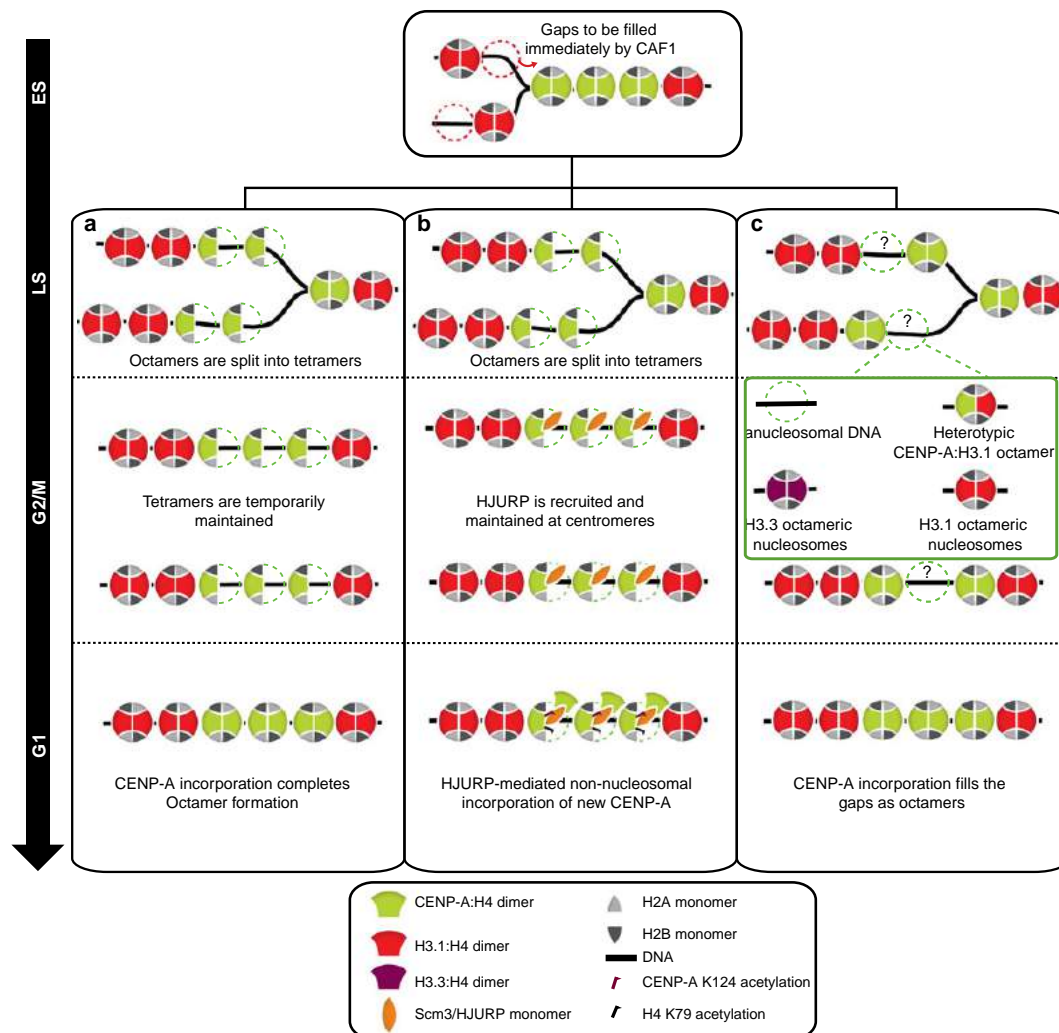


Figure 9. Modèle de la structure dynamique des nucléosomes centromériques au cours du cycle cellulaire. a. Après le passage de la fourche de réplication en phase S, les octamères d’histones contenant CenH3 sont partagés en deux tétramères sur chaque brin d’ADN fils permettant l’hérité fidèle de la marque épigénétique. De nouvelles molécules de CenH3 sont incorporées pendant la phase G1. **b.** Même modèle mais l’histone chaperone HJURP/Smc3 participe à la stabilisation des tétramères. **c.** Modèle alternatif dans lequel les nucléosomes centromériques restent sous forme d’octamères tout au long du cycle. Pendant la réplication, les nucléosomes sont répartis aléatoirement entre les deux brins d’ADN. Les trous générés par le passage de la fourche de réplication pourraient rester vides ou être comblés temporairement par l’incorporation d’autres nucléosomes (nucléosomes canoniques ou contenant le variant H3.3).

Padeganeh et al., Chromosome Res. 2013

L'incorporation des histones dans la chromatine dépend de protéines chaperones. Chez l'Homme, les facteurs RbAp46/48 et HJURP (Smc3 chez *cerevisiae*) sont nécessaires pour empêcher la dégradation des molécules solubles de CenH3 et ont donc été considérées comme des chaperones de CenH3 (Dunleavy et al., 2009; Foltz et al., 2009; Hayashi et al., 2004). RbAp46/48 semble être une chaperonne non spécifique par conséquent, l'essentiel des recherches se sont concentrées sur HJURP. Cette dernière présente une activité d'assemblage des nucléosomes spécifiquement pour les molécules de Cenp-A nouvellement synthétisées (Barnhart et al., 2011). Le domaine de fixation de Cenp-A dans la région N-terminale de HJURP reconnaît le domaine CATD de Cenp-A (Shuaib et al., 2010). HJURP pourrait également interagir avec la région N-terminale de Cenp-A pour stabiliser l'association entre Cenp-A et H4 (Bassett et al., 2012). L'assemblage des nucléosomes canoniques dans la chromatine est bien décrit et consiste en l'incorporation initiale d'un tétramère [H3-H4]₂ dans l'ADN, puis l'addition de deux dimères H2A-H2B complète le nucléosome (Smith and Stillman, 1991). Concernant les nucléosomes CenH3, des études ont montré que HJURP fixait des dimères Cenp-A-H4 (Hu et al., 2011). Il n'est donc toujours pas compris si Cenp-A est incorporé dans la chromatine sous forme de dimère ou de tétramère. La localisation de HJURP au niveau des centromères dépend du complexe de « licensing » Mis18 mais aucune interaction physique entre les deux facteurs n'a encore été identifiée.

3. Rôle de la chromatine centromérique

Chez la plupart des organismes, les régions centromériques contiennent des nucléosomes canoniques intercalés entre les nucléosomes CenH3. Chez la Drosophile, l'Homme et les plantes, ces nucléosomes sont diméthylés sur la lysine 4 (H3K4me2), une marque caractéristique de l'activation transcriptionnelle de la chromatine (Sullivan and Karpen, 2004). Cette modification est importante pour la régulation de l'assemblage des centromères. En effet, la perte de cette marque altère le recrutement de HJURP aux centromères et empêche l'incorporation de molécules de Cenp-A nouvellement synthétisées au niveau de l'ADN alpha satellite d'un chromosome artificiel humain (HAC). La perte de H3K4me2 conduit également à la diminution de la transcription de cet ADN (Bergmann et al., 2011). Etant donné le rôle de la méthylation des histones dans la structure de la chromatine, il est probable que la topologie de la chromatine joue un rôle déterminant dans la localisation de CenH3. De manière cohérente avec la précédente

étude, il a été montré que le ciblage d'une lysine méthyltransférase ciblant H3K9 (Suv39h1) au niveau d'un HAC empêchait l'assemblage *de novo* de Cenp-A dans l'ADN de ce HAC. A l'inverse, le ciblage d'histone acétyltransférases augmentant le niveau d'acétylation de H3K9 accroît la quantité de Cenp-A sur l'ADN satellite. Cette modification est seulement nécessaire pour l'assemblage *de novo* de Cenp-A car l'élimination des acétyltransférases n'a pas d'effet sur les centromères préexistants (Ohzeki et al., 2012).

4. Un rôle pour la transcription ?

La balance entre la méthylation et l'acétylation de H3K9 est généralement liée à l'activité transcriptionnelle indiquant un possible lien entre la transcription des régions centromériques et la localisation de CenH3. En accord avec cette hypothèse, des transcrits homologues à l'ADN centromérique ont été détectés chez les mammifères et chez les plantes, ces derniers sont maturés en siARN chez les plantes (Bouzinba-Segard et al., 2006; May et al., 2005; Neumann et al., 2007). Des gènes actifs, peuvent résider dans la chromatine centromérique et l'ARN polymérase II et ses facteurs de transcription associés ont également été observés au niveau des centromères (Saffery et al., 2003; Yan et al., 2006). Les transcrits centromériques pourrait faciliter le recrutement des facteurs requis pour la déposition des protéines CenH3, à la manière dont les siARN ou autres ARN non codants recrutent des enzymes de modification des histones. Il a également été proposé que les transcrits centromériques naissant puissent s'hybrider avec leur matrice pour former un hybride ADN-ARN qui pourrait déclencher une voie de réponse aux dommages de l'ADN. Cela conduirait à un remodelage de la chromatine qui permettrait l'incorporation de CenH3 (Allshire and Karpen, 2008). Cependant, si la transcription semble jouer un rôle dans l'établissement ou le maintien de la chromatine centromérique, il reste à déterminer si c'est le processus de transcription ou si ce sont les transcrits eux-mêmes qui sont impliqués.

IV. Centromères : quand stabilité rime avec plasticité

Les centromères sont des éléments essentiels qui présentent généralement une localisation et une fonction stable au sein des populations cellulaires d'un organisme. Pourtant, en raison de leur nature épigénétique, les centromères sont également pourvus d'une étonnante plasticité. Certaines séquences non centromériques peuvent occasionnellement acquérir et maintenir une fonction centromérique, formant un néocentromère. A l'inverse, certaines régions contenant des centromères fonctionnels peuvent être inactivées et perdre leur fonction initiale. Ces phénomènes se produisent généralement en réponse à des réarrangements génomiques qui altèrent la structure des chromosomes et perturbent le nombre de centromères présents sur chaque chromosome (Sullivan et al., 2001).

1. Chromosomes dicentriques

Chez les espèces aux chromosomes monocentriques, le nombre de centromères présents sur chaque chromosome est régulé de façon à maintenir un et un seul centromère actif sur chaque chromosome. Des réarrangements génomiques tels que des translocations ou des fusions de télomères peuvent néanmoins conduire à la formation de chromosomes dicentriques sur lesquels deux centromères actifs sont physiquement liés. Le destin des chromosomes dicentriques a été décrit pour la première fois chez le maïs par Barbara McClintock qui montra que la présence de deux centromères sur un même chromosome était une situation instable, dans laquelle les deux centromères avaient tendance à s'orienter vers des pôles opposés conduisant à la formation de ponts et de cassures chromosomiques pendant la division cellulaire (McClintock, 1939, 1941). Ceci a ensuite été observé pour les chromosomes dicentriques de la *Drosophile* et de la levure *S. cerevisiae* (Hill and Bloom, 1989).

Cependant, chez les mammifères, les plantes et la levure *S. Pombe*, les chromosomes dicentriques peuvent cependant être stabilisés et hérités normalement au cours des divisions cellulaires. Cette stabilité est attribuable à un phénomène d'inactivation de l'un des deux centromères (Earnshaw and Migeon, 1985; Han et al., 2006; Sato et al., 2012). Le chromosome dicentrique se comporte alors comme un chromosome monocentrique capable de ségréger classiquement. L'inactivation

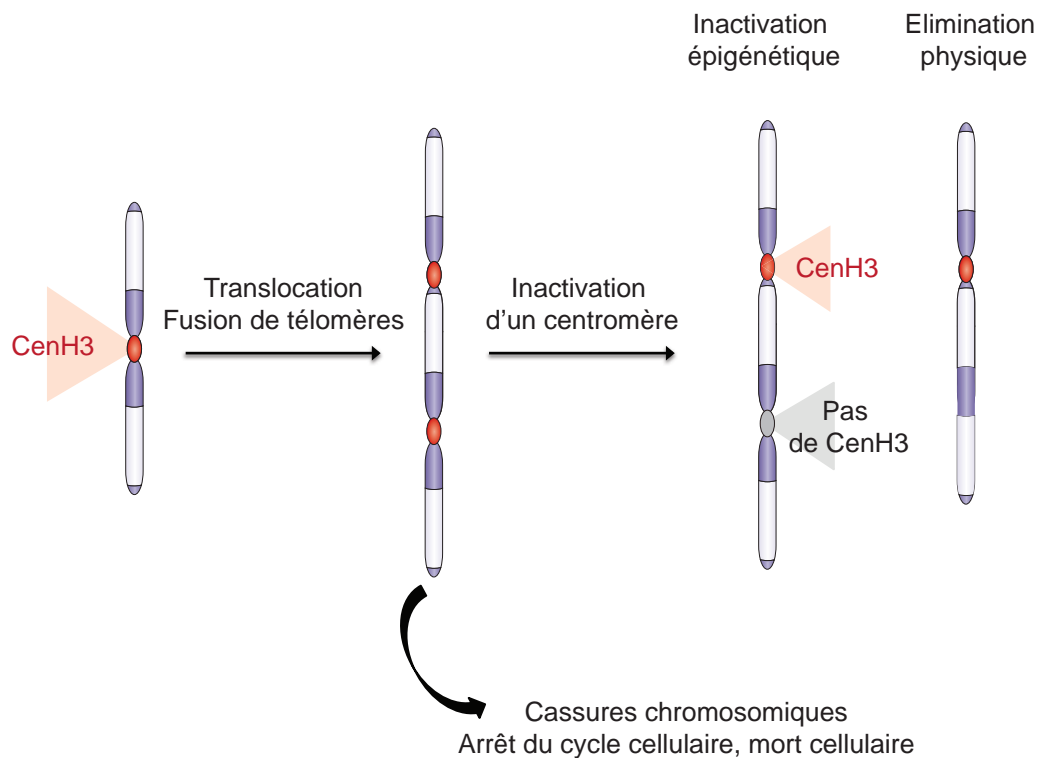


Figure 10. Mécanismes d'inactivation des centromères. Des réarrangements génomiques (translocations, fusions de télomères) peuvent conduire à la formation de chromosomes dicentriques. Dans la plupart des cas, les chromosomes dicentriques sont instables et subissent des cassures chromosomiques en mitoses qui conduisent à l'arrêt du cycle cellulaire et à la mort cellulaire. Certains chromosomes dicentriques peuvent cependant être stabilisés par l'inactivation d'un des deux centromères. L'inactivation peut s'effectuer sans altération de la séquence d'ADN centromérique (inactivation épigénétique) ou par l'élimination physique des séquences centromériques.

des centromères se traduit moléculairement par la disparition des protéines du kinétochore et notamment des protéines CenH3 (Earnshaw and Migeon, 1985; Sullivan and Schwartz, 1995).

Certains chromosomes peuvent également rester dicentriques et fonctionnels (Higgins et al., 2005; Page and Shaffer, 1998; Stimpson et al., 2010). Il a été suggéré que la distance inter-centromérique pourrait être corrélée au destin des chromosomes dicentriques. En effet, chez la levure et l'Homme, les centromères les plus distants sont plus fréquemment inactivés que les centromères physiquement proches. (Koshland et al., 1987; Page and Shaffer, 1998; Sullivan and Willard, 1998).

2. Stabilisation des chromosomes dicentriques par inactivation des centromères

Deux types de centromères inactivés ont été reportés : des centromères inactifs mais contenant des séquences d'ADN intactes et des centromères inactifs présentant des délétions de l'ADN centromérique. Dans le premier cas, l'inactivation des centromères est qualifiée d'épigénétique car elle altère la fonction de la région chromatinienne de manière héritable sans affecter la séquence des centromères. La deuxième situation correspond à une inactivation génomique, par élimination physique des séquences centromériques. Dans les deux cas, l'inactivation s'accompagne de la disparition du marqueur centromérique CenH3 (Stimpson and Sullivan, 2010) (Figure 10). Actuellement, les mécanismes moléculaires impliqués dans les deux types d'inactivation sont encore très peu compris, en particulier parce que le phénomène d'inactivation des centromères est rare chez les modèles étudiés et parce qu'il est difficile de produire artificiellement des chromosomes dicentriques.

Une étude récente menée chez *S. pombe* a cependant permis la création de chromosomes dicentriques artificiels en utilisant des systèmes de recombinaison site-spécifique ou de recombinaison méiotique entre les chromosomes II et III (Sato et al., 2012). Chez cette levure, la présence d'un chromosome dicentrique est létale pour la majorité des cellules (99% des cellules). Bien que la majorité des cellules meure, une petite proportion de chromosome dicentriques est maintenue de manière stable (1%). Cette stabilisation s'est établie de trois façons différentes : i. par un événement de cassure entre les deux centromères permettant un retour à un karyotype normal (10%), ii. Par la délétion physique de l'un des deux centromères sur le chromosome fusionné (10%), iii. Par un processus d'inactivation épigénétique d'un des centromères qui n'a

altéré ni la structure du chromosome dicentrique ni la séquence d'ADN des centromères (80%). Dans ce dernier cas, la région centrale du centromère inactivé ne contient plus Cnp1 (CenH3) et porte des signatures de l'hétérochromatine avec un enrichissement en H3K9me2 et un appauvrissement des marques H3K9ac et H3K14ac. L'hétérochromatine n'est pas nécessaire pour l'initiation de l'inactivation mais sert au maintien de l'état inactivé.

Chez les plantes, des chromosomes dicentriques stables ont été décrits chez le maïs et le blé (Gao et al., 2011; Han et al., 2006; Zhang et al., 2010). Les cellules de maïs contiennent des petits chromosomes extra-numéraires (chromosomes B) qui ne sont pas indispensables. Des expériences d'hybridation *in situ* utilisant une sonde dirigée contre les répétitions centromériques couplées à des immunofluorescences avec un anticorps CenH3 ont montré que certains minichromosomes portaient deux centromères B dont seulement un est actif (Han et al., 2006). Des études complémentaires impliquant des chromosomes dicentriques avec des tailles différentes de centromères B ont montré que le plus petit centromère était préférentiellement inactivé. Une fois séparés par des recombinaisons intra-chromosomiques, le petit centromère peut être réactivé en récupérant les protéines CenH3 et Cenp-C. L'inactivation et la réactivation de ces centromères se produit sans altération des séquences d'ADN (Han et al., 2009). Comme chez *S. pombe*, les centromères de plantes inactivés adoptent une structure hétérochromatinienne. En effet, la méthylation des cytosines a été cartographiée dans les séquences répétées de cellules de maïs, révélant que les centromères actifs sont hypométhylés tandis que les centromères inactifs sont hyperméthylés (Koo et al., 2011). De manière similaire, chez le blé, les centromères inactivés épigénétiquement semblent contenir un niveau plus élevé de H3K27me2 et H3K27me3, signes d'une forme d'hétérochromatinisation de la région centromérique (Zhang et al., 2010).

Chez l'Homme, les translocations robertsonniennes impliquant les chromosomes acrocentriques 13, 14, 15, 21 et 22 génèrent des chromosomes dicentriques qui sont maintenus dans la population par l'inactivation d'un des centromères. Les centromères inactivés ont perdu leur association avec Cenp-A, Cenp-C et Cenp-E (Sullivan and Schwartz, 1995). Sur ces chromosomes dicentriques, certains centromères, comme celui du chromosome 14, sont moins souvent inactivés que d'autres comme s'il existait des centromères plus « forts » que d'autres (Sullivan et al., 1994a).

Un chromosome dicentrique artificiel a été créé avec deux bras courts du chromosome X humain dans des cellules hybrides souris/Homme (Higgins et al., 1999). Différentes issues ont été

observées : 1. le chromosome est devenu fonctionnellement monocentrique grâce à l'inactivation de l'un des deux centromères, 2. le chromosome est resté fonctionnellement dicentrique, 3. L'activité du centromère est hétérogène : l'état monocentrique ou dicentrique peut changer au sein d'une même population de cellules en croissance. Des analyses moléculaires ont montré que la taille de l'ADN alpha satellite au niveau des centromères inactivés n'avait pas changé suggérant que l'inactivation s'est produite par un mécanisme épigénétique. Dans cette étude, la décision d'inactivation d'un des centromères semble indépendante de la taille des espaces inter-centromériques (Higgins et al., 2005).

Plus récemment, des chromosomes dicentriques ont été générés dans des cellules humaines exprimant une version mutante de la protéine associée aux télomères TRF2 qui génère des fusions de télomères (Stimpson et al., 2010). Cette étude a révélé qu'une proportion importante des chromosomes dicentriques maintenait deux centromères actifs de manière stable sur plus de 20 générations. Les autres chromosomes subissent l'inactivation de l'un des centromères entre 4 jours et 20 semaines après la formation de leur création. Dans la moitié de ces cas, l'inactivation des centromères s'accompagne de l'apparition transitoire de petits fragments de chromosomes qui contiennent, sur la base d'expériences de FISH et d'immunofluorescence, de l'ADN alpha satellite et des protéines centromériques. De plus, une analyse par FISH semi-quantitatif a montré que la taille de l'ADN satellite était réduite après le processus d'inactivation du centromère. Ces résultats ont donc suggéré qu'un des processus d'inactivation des centromères chez l'Homme implique la délétion partielle de séquences centromériques. Comme les centromères sont constitués de nucléosomes CenH3 intercalés dans des nucléosomes H3, l'inactivation des centromères pourrait s'expliquer par l'élimination physique des séquences d'ADN qui portent les nucléosomes CenH3.

D'une manière générale, il semble que l'apparition de chromosomes dicentriques dans les génomes eucaryotes puisse avoir des issues différentes en fonction des organismes et des types cellulaires. La capacité des cellules à inactiver l'un des centromères d'un chromosome dicentrique suggère qu'il existe un mécanisme « senseur » du nombre de centromères actifs sur chacun des chromosomes. L'inactivation des centromères implique également des mécanismes variables d'un organisme à l'autre qui altèrent complètement, partiellement ou n'altère pas du tout la séquence de l'ADN centromérique. Les bases moléculaires de ces mécanismes sont, pour l'instant, incomprises. L'inactivation épigénétique pourrait utiliser des systèmes d'éviction des nucléosomes

CenH3 de l'ADN centromérique ainsi que des mécanismes bloquant l'incorporation de CenH3 pendant la réplication. L'inactivation par élimination physique des séquences centromériques pourrait quant à elle nécessiter des processus de recombinaison, de cassures de l'ADN induites par la réplication et des cassures impliquant la voie du NHEJ (Pennaneach and Kolodner, 2009; Stimpson et al., 2012).

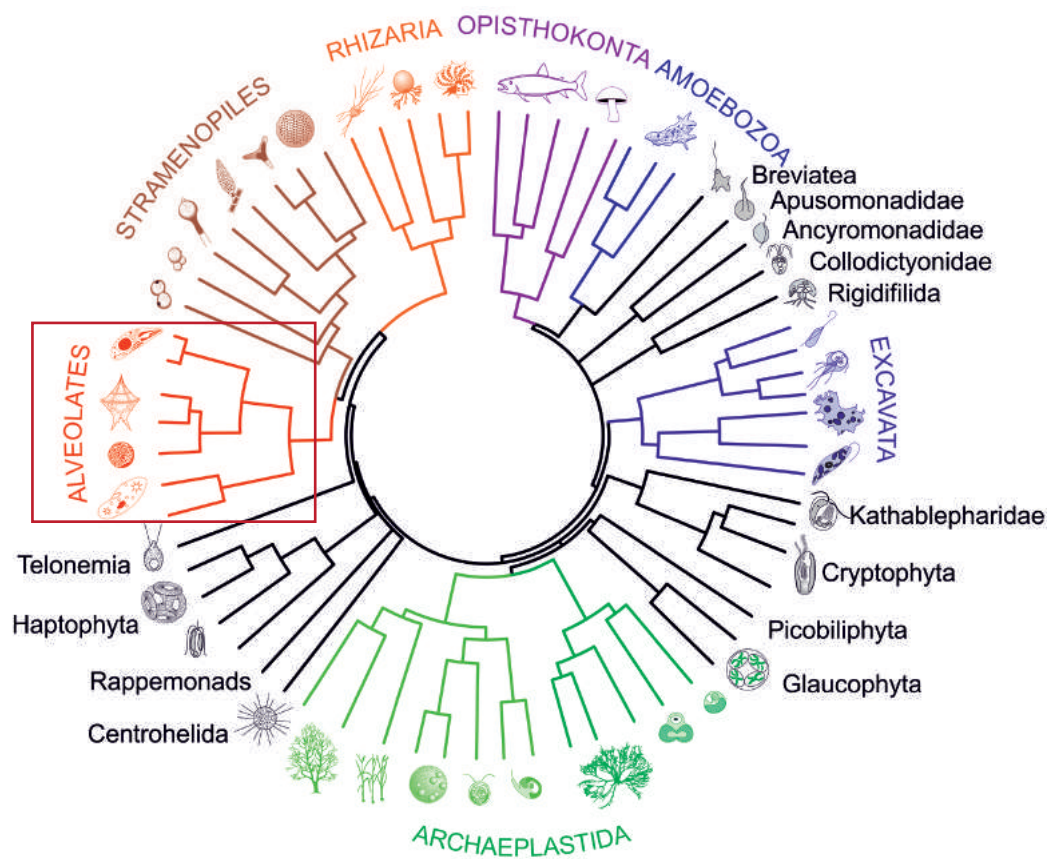


Figure 11. Phylogénie des eucaryotes. Le groupe des Alvéolés auquel appartient *Paramecium tetraurelia* est encadré en rouge.
Pawlowski, *BMC Biology*, 2013

CHAPITRE 3

***PARAMECIUM TETRAURELIA* : Organisme modèle pour l'étude de la régulation épigénétique de l'organisation du génome**

I. Dimorphisme nucléaire et cycle de vie

Paramecium tetraurelia est un eucaryote unicellulaire non pathogène, très répandu dans les eaux douces, qui se nourrit de bactéries et qui peut facilement être cultivé en laboratoire. Il appartient au phylum des ciliés dans le groupe des organismes alvéolés, un groupe distinct des plantes (Archaeplastida), des animaux et des champignons (Opisthokonta) (Pawlowski, 2013) (Figure 11). Pour différentes raisons qui vont être présentées par la suite, *P. tetraurelia* s'avère être un modèle génétique particulièrement adapté pour l'étude des mécanismes qui contrôlent l'organisation chromatinienne des génomes et qui régulent l'activité des centromères.

Chez *P. tetraurelia*, comme chez tous les ciliés, les fonctions somatique et germinale sont séparées dans deux types de noyaux distincts qui partagent le même cytoplasme. Deux micronoyaux, diploïdes, assurent les fonctions germinales. Ils ne sont pas exprimés pendant la croissance végétative des cellules mais sont responsables de la transmission du matériel génétique à chaque cycle de reproduction sexuée. Un macronoyau, hautement polyploïde (environ 800n) assure, quant à lui, les fonctions somatiques en prenant en charge la transcription des gènes tout au long du cycle de vie. En contrepartie, le macronoyau n'est pas transmis à la génération sexuée suivante (Chalker et al., 2013).

En conditions de températures adaptées (14-30°C), les cellules de paramécies peuvent se reproduire de façon asexuée (croissance végétative) par division clonale. L'épuisement des ressources nutritives conduit au déclenchement du cycle de reproduction sexuée qui peut se produire de deux façons différentes, soit par conjugaison entre deux cellules suffisamment jeunes et de types sexuels opposés, soit par autogamie, un processus d'autofécondation qui se produit généralement chez les cellules plus âgées. Le cycle sexuel débute par l'entrée en méiose des deux micronoyaux qui vont générer huit noyaux haploïdes. Sept vont dégénérer et un va subir une

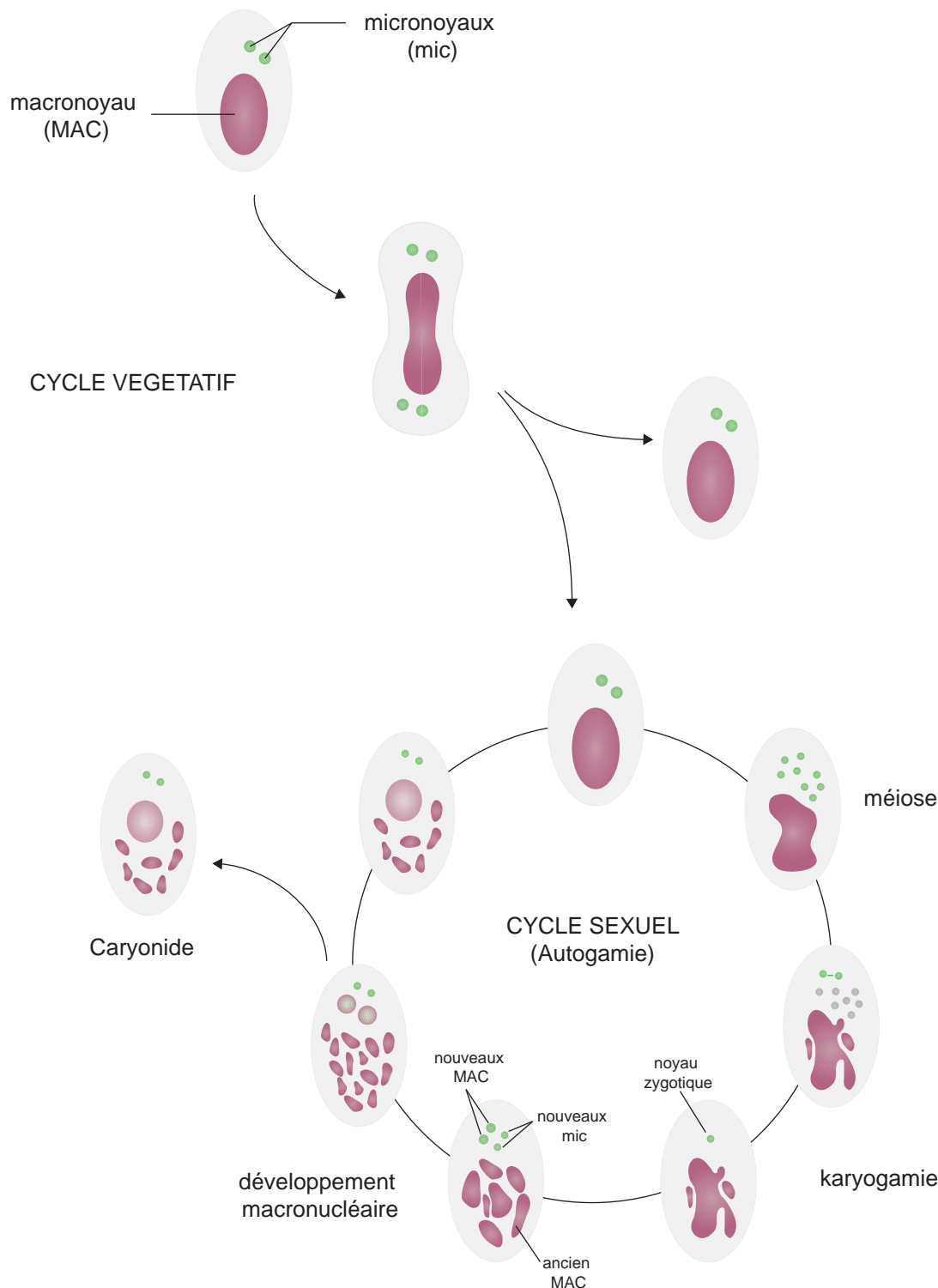


Figure 12. Cycle de vie de *Paramecium tetraurelia*. En croissance végétative, les cellules de Paramecies se multiplient par division clonale. Après épuisement des ressources nutritives, les cellules entrent dans le cycle sexuel soit par autogamie (autofécondation), soit par conjugaison. Le déroulement de l'autogamie est présenté sur ce schéma. Les micronoyaux subissent une méiose tandis que le macronoyau commence à se dégrader. Seul un produit de méiose est maintenu et va subir une mitose pour générer deux noyaux haploïdes identiques. Ces derniers fusionnent pour former le noyau zygotique, complètement homozygote. Le noyau zygotique se divise deux fois et les produits de mitoses vont se différencier en deux nouveaux micronoyaux et deux nouveaux macronoyaux. La cellule se divise ensuite pour donner deux caryonides contenant chacun deux micronoyaux et un macronoyau.

division mitotique donnant lieu à deux noyaux gamétiques haploïdes génétiquement identiques. Dans le cas de l'autogamie, ces deux produits de mitose fusionnent pour former le noyau zygotique, diploïde et par conséquent complètement homozygote. Dans le cas de la conjugaison, l'un des deux noyaux est échangé réciproquement entre les deux cellules conjugantes avant de fusionner avec le produit haploïde de la cellule d'origine et de générer le noyau zygotique diploïde. Après deux mitoses successives du noyau zygotique, les cellules conjugantes se séparent et les quatre noyaux diploïdes vont se différencier en deux nouveaux micronoyaux, aux futures fonctions germinales et deux nouveaux macronoyaux, aux futures fonctions somatiques. La différenciation des noyaux germinaux et somatiques s'effectue au cours d'une période d'une dizaine d'heures qualifiée de développement macronucléaire. Suite au développement, la cellule se divise, les micronoyaux subissent une mitose et les deux nouveaux macronoyaux sont répartis dans chacune des cellules filles appelées caryonides. Au cours du cycle de reproduction, l'ancien macronoyau, que l'on nomme généralement macronoyau maternel, est progressivement dégradé mais reste dans la cellule sous forme de fragments. Il contribue à l'expression génique jusqu'à ce que les nouveaux macronoyaux soient fonctionnels. Les fragments de l'ancien macronoyau sont ensuite dégradés et dilués au cours des divisions végétatives qui suivent le cycle sexuel (Figure 12) (Chalker et al., 2013).

II. Génome...

Avec environ 40 000 gènes annotés codant pour des protéines, *P. tetraurelia* contient l'un des génomes les plus riches en gènes connu actuellement. Ce nombre très élevé de gènes, presque deux fois plus important que chez les mammifères, est la conséquence de trois duplications successives du génome (WGD : Whole Genome Duplication). Les paires de chromosomes dupliqués ont progressivement divergé par des mutations ponctuelles et des délétions. De nombreux gènes ont été perdus ou sont devenus des pseudogènes mais les gènes très fortement exprimés ont été fréquemment retenus en plusieurs copies (Aury et al., 2006).

Micronoyaux et macronoyau se distinguent par leur contenu génétique. Tandis que les deux micronoyaux contiennent la version complète du génome de *Paramecium* dont la taille est estimée à une centaine de Mb, le macronoyau contient une version polyploïde et réarrangée, dans laquelle environ 30 Mb d'ADN ont été éliminés. Le génome macronucléaire correspond à 800 copies de

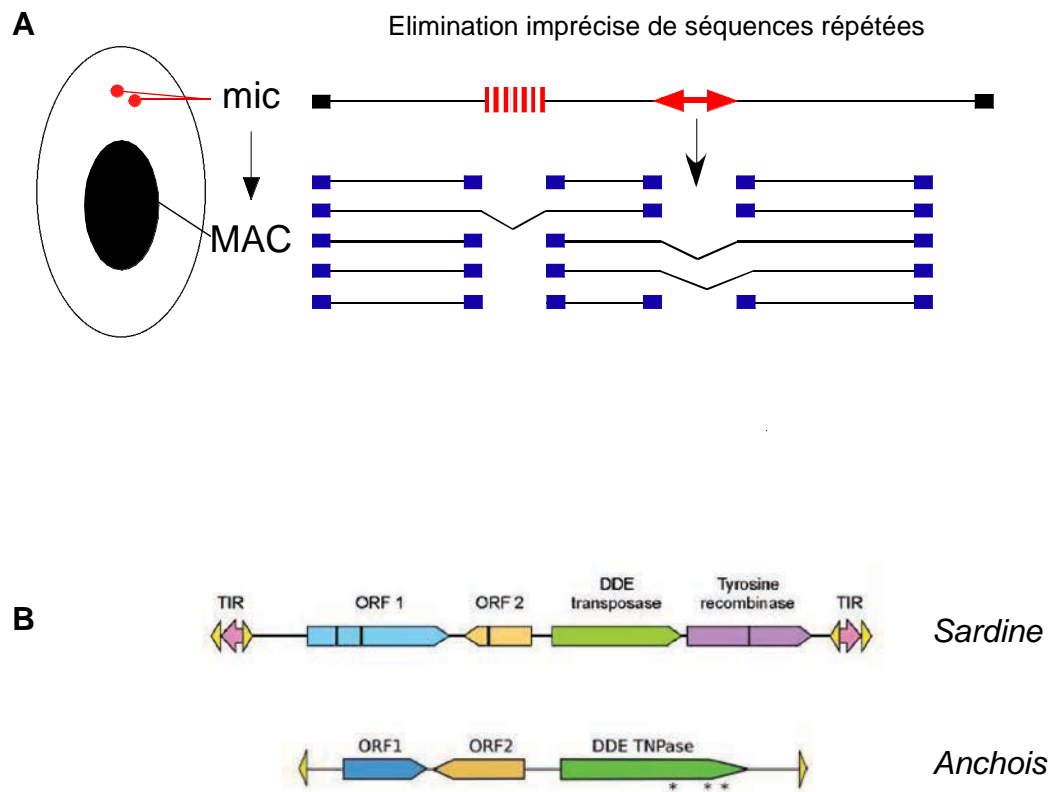


Figure 13. Elimination imprécise de séquences répétées au cours du développement macronucléaire de *P. tetraurelia*. **A.** Des régions minisatellites (barres verticales rouges) et des transposons (flèche rouge) sont imprécisément éliminés du génome micronucléaire pour former le génome macronucléaire qui est par ailleurs endorépliqué. Ces événements d'élimination conduisent à des délétions et à la fragmentation des chromosomes macronucléaires. **B.** Transposons de la famille des Tc1/Mariner retrouvés dans le génome de la Paramécie. Représentation schématique des séquences consensus des transposons *Sardine* et *Anchois* montrant la position des ORF.

Duharcourt et al., Trends in Genet., 2009

Arnaiz et al., PloS Genet., 2012

72 Mb d'ADN très compacté, contenant 78% de séquences codantes séparées par des régions intergéniques de 352 pb en moyenne (Aury et al., 2006; Zagulski et al., 2004). C'est au cours du développement macronucléaire, après les événements sexuels, que le génome est endorépliqué et que des séquences micronucléaires sont éliminées de manière massive et reproductible pour façonner le génome macronucléaire qui sera transcriptionnellement actif.

III. Réarrangements du génome associés au développement macronucléaire

1. Endoréplication

Le génome micronucléaire, présent en deux copies dans le noyau zygotique, est fortement endorépliqué pour atteindre environ 800 copies dans les macronoyaux en développement. Peu de choses sont connues sur les mécanismes de réplication mis en jeu et sur le timing des cycles de réplication au cours du développement macronucléaire. Il a cependant été montré que quelques cycles d'endoréplication avaient lieu avant les événements d'élimination d'ADN (présentés plus bas). 16 à 32 copies de chaque chromosome germinale pourraient être présentes quand les premières cassures sont introduites (Bétermier et al., 2000; Ching and Berger, 1986).

2. Elimination imprécise de séquences répétées

Des séquences répétées, notamment des transposons et des régions minisatellites, sont éliminées du génome micronucléaire sans précision du site de coupure. Ce phénomène conduit d'une part, à des délétions jusqu'à de plusieurs dizaines de kilobases dont les bornes peuvent être variables sur les différentes copies de chaque chromosome, et d'autre part à la fragmentation des chromosomes micronucléaires après l'ajout de télomères à l'extrémité des sites de coupures (Le Mouel et al., 2003) (Figure 13A). Le génome micronucléaire n'étant pas encore complètement assemblé, ces régions imprécisément éliminées sont encore peu connues mais on estime leur complexité à environ 25 Mb (Arnaiz et al., 2012). Trois familles de transposons appartenant à la superfamille des Tc1/Mariner ont cependant pu être identifiées parmi les banques d'ADN micronucléaires disponibles (Steele et al., 1994). Les transposons appelés *Thon* et *Sardine* ont de longs TIR (Terminal Inverted Repeat) palindromiques, une tyrosine recombinase et une transposase de type

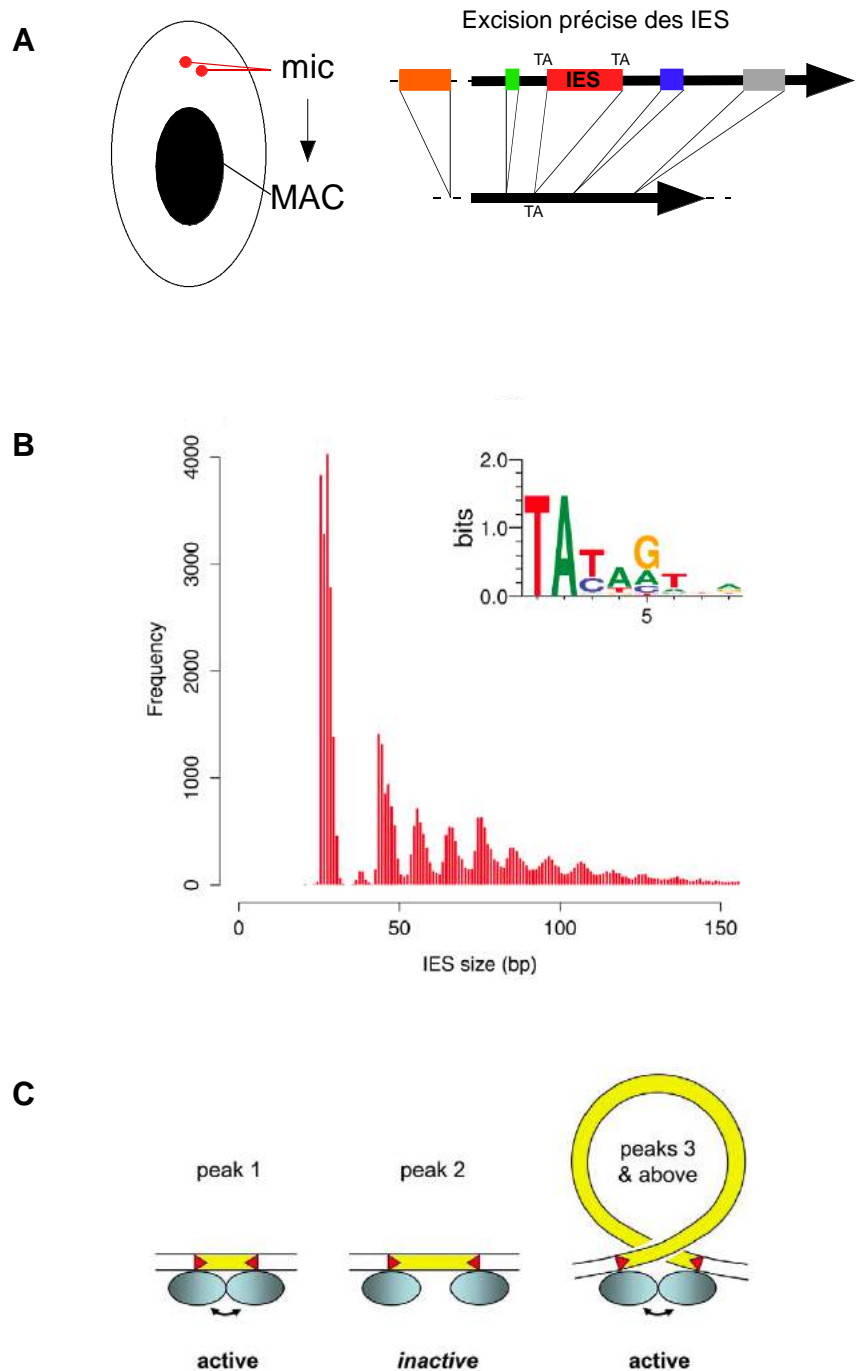


Figure 14. Elimination précise des IES. **A.** Les IES, des séquences courtes et uniques, sont éliminées très précisément du génome macronucléaire. Leur excision permet de rétablir les phases ouvertes de lecture des gènes. **B.** Distribution de taille des IES inférieures à 150 pb. Une séquence logo montrant la composition des extrémités des IES est présentée en haut à droite du diagramme. **C.** Modèle présentant les mécanismes d'assemblage du complexe d'excision en fonction de la taille des IES. Au niveau des IES les plus courtes (premier pic : 26-32 pb), les complexes présents aux deux extrémités des IES peuvent interagir directement et être activés. A partir du troisième pic (42-52 pb), l'interaction entre les deux complexes d'excision s'établit grâce à la formation d'une boucle d'ADN double brin. Le deuxième pic (32-42 pb) est très peu représenté car les IES de cette taille sont à la fois trop courtes et trop longues pour qu'un complexe d'excision actif se forme.

DDE caractéristique de la sous-famille IS630/Tc1. Les transposons *Anchois* sont quant à eux caractérisés par des TIR plus courts, qui ne contiennent pas de palindrome interne, et par l'absence de tyrosine recombinase (Arnaiz et al., 2012) (Figure 13B).

3. Elimination précise des IES

Environ 45 000 séquences uniques, appelées IES (Internal ExcisEd Sequences), dispersées sur l'ensemble du génome micronucléaire sont excisées de façon très précise au nucléotide près (Gratias and Bétermier, 2001). Les IES représentent 3,55 Mb de la complexité du génome micronucléaire (Arnaiz et al., 2012). Elles sont présentes dans les régions géniques comme dans les régions intergéniques et leur excision est requise pour rétablir la phase ouverte de lecture de nombreux gènes qui seront exprimés par le génome macronucléaire (Figure 14A). La précision des sites de coupures pour l'excision des IES est donc indispensable à la fonctionnalité du macronoyau et à la vie des cellules. Toutes les IES sont bornées à leurs extrémités par un dinucléotide TA dont une copie est retrouvée au niveau de la jonction macronucléaire après l'excision (Bétermier, 2004; Klobutcher and Herrick, 1995). Cette caractéristique est nécessaire à l'excision des IES puisque une seule substitution de base dans ce dinucléotide rend l'excision de l'IES concernée défectueuse (Mayer and Forney, 1999). De plus, un consensus faiblement conservé, commençant par le TA et s'étendant dans l'IES (5'-TAYAGYNR-3'), peut être défini à l'extrémité d'une majorité d'IES. De manière remarquable, cette séquence ressemble au consensus trouvé aux extrémités des transposons de la superfamille des Tc1/Mariner (Klobutcher and Herrick, 1995). Cette correspondance pourrait attester que, d'un point de vue évolutif, au moins une partie des IES sont des restes de transposons Tc1/Mariner dont les séquences auraient fortement dégénéré, sauf aux extrémités. De plus, les éléments transposables *Anchois* sont en réalité des IES (ils sont excisés précisément) ce qui soutient le lien entre IES et transposons (Arnaiz et al., 2012). Un faible consensus aurait été conservé aux extrémités car il serait finalement devenu nécessaire à l'efficacité de la machinerie d'excision (Klobutcher and Herrick, 1997).

La taille des IES varie entre 26 pb et 5 kb mais la grande majorité d'entre elles sont courtes : 93% des IES sont plus petites que 150 pb et 35% des IES ont une taille comprise entre 26 et 32 pb (Arnaiz et al., 2012). La distribution de taille de l'ensemble des IES forme de manière assez

spectaculaire une courbe sinusoïdale de période 10 pb qui pourrait refléter une contrainte physique associée au processus d'excision (Figure 14B). Il a été suggéré que la structure de la double hélice d'ADN dont le tour d'hélice fait environ 10 pb pourrait jouer un rôle dans l'accès de la machinerie d'excision aux séquences éliminées. Les IES auraient été sélectionnées pour avoir une taille permettant de présenter les deux complexes de la machinerie d'excision sur la même face de la double hélice. A partir du 3^{ème} pic (42-52 pb), les IES seraient excisées par un mécanisme impliquant la formation d'une boucle d'ADN double brin qui faciliterait l'interaction entre les deux complexes d'excision présents aux extrémités des IES. De manière cohérente avec cette hypothèse, des produits d'excision d'IES circulaires ont été détectés pour les IES les plus grandes (Bétermier et al., 2000). Les IES du premier pic (26-32 pb) seraient suffisamment courtes pour être intégrées dans un complexe d'excision sans formation de boucle d'ADN. En revanche, très peu d'IES sont présentes dans le deuxième pic (32-42 pb). Celui-ci représenterait une taille « interdite » car les IES de cette longueur seraient trop courtes pour former des boucles et trop longues pour permettre l'interaction des complexes d'excision aux extrémités de chaque IES (Figure 14C). Éliminées avec une beaucoup moins bonne efficacité, ces IES auraient été contre-sélectionnées au cours de l'évolution (Arnaiz et al., 2012).

4. L'endonucléase domestiquée PiggyMac

Les IES étant des restes de transposons Tc1/Mariner, il a été envisagé que l'enzyme responsable de leur excision pourrait être une transposase retenue dans une ou plusieurs copies de transposons intacts et toujours actives. Les transposases Tc1/Mariner mobilisent leurs éléments d'ADN par un mécanisme de « couper-coller » initié par un clivage flanquant un dinucléotide TA et laissant 2 bases débordantes en 3' (Plasterk et al., 1999). Pourtant, il a été observé que les extrémités libres générées après l'excision des IES étaient débordantes de 4 bases en 5' et centrées autour du TA, impliquant que la coupure devait être produite par une endonucléase d'une autre famille (Gratias and Bétermier, 2003). *In vitro*, les transposases Piggybac catalysent des cassures double brin aux extrémités des transposons Piggybac générant 4 bases débordantes en 5' (Mitra et al., 2008). Cette similitude avec le mécanisme d'excision des IES a permis d'identifier dans le génome de la *Paramecia* l'endonucléase responsable des cassures double brin aux bornes des éléments éliminés. Celle-ci a été nommée en conséquence PiggyMac pour son rôle dans le développement macronucléaire (Baudry et al., 2009). Le gène codant pour PiggyMac est présent en copie unique

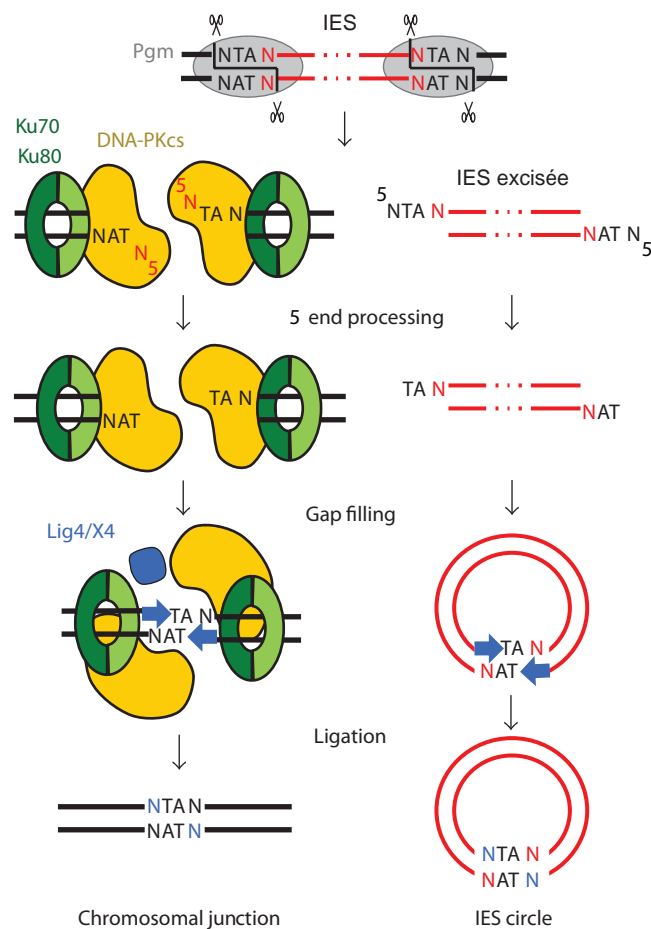


Figure 15. Mécanismes moléculaires impliqués dans l'excision des IES chez *Paramecium*. La première étape correspond à l'introduction de cassures d'ADN double brin décalées de 4 bases par l'endonuclease domestiquée PiggyMac à chaque extrémité des IES. Selon la voie connue du NHEJ, les protéines Ku70/Ku80 et DNA-PKcs pourraient être recrutées au niveau des extrémités cassées. Le nucléotide en 5' est éliminé, les protéines impliquées dans ce processus n'ont pas encore été identifiées. Les dinucléotides TA centraux présents à chaque extrémité cassée pourraient s'apparier pour faciliter la réparation. Le complexe ligase IV/XRCC4 est requis pour le recrutement ou l'activation d'une ADN polymérase qui va resynthétiser les nucléotides manquants avant la ligation finale. Les IES excisées pourraient être circularisées par un mécanisme similaire quand elles sont suffisamment longues.

Dubois et al, Int J Evol Biol., 2012

dans le génome et n'est plus inclus dans un transposon indiquant que l'endonucléase a été domestiquée. La protéine PiggyMac contient un domaine catalytique conservé contenant une série de résidus Asparagine (domaine DDD), un domaine PHD (Plant Homéodomain) potentiellement impliqué dans la reconnaissance de la chromatine et un domaine coiled-coil qui pourrait lui permettre d'interagir avec d'autres facteurs protéiques. La comparaison du génome de cellules inactivées pour PiggyMac avec le séquençage de banques d'ADN micronucléaire a confirmé que PiggyMac était requise pour l'élimination de toutes les séquences micronucléaires connues (IES et régions imprécisément éliminées). Ce génome a récemment permis l'identification et la caractérisation des 45 000 IES décrites ci-dessus et aidera dans l'avenir à l'assemblage des régions répétées imprécisément éliminées (Arnaiz et al., 2012).

5. Réparation des cassures double brin

Le complexe Ligase IV-XRCC4, un facteur essentiel de la voie du NHEJ (Non-Homologous End Joining), est requis pour la réparation des cassures double brin au niveau des IES (Kapusta et al., 2011). Il a été proposé qu'un couplage entre excision et réparation favorise la précision de la réparation. En effet, les extrémités cassées pourraient être maintenues associées par le dinucléotide TA central jusqu'à ce que la résection, la réparation et la ligation des deux extrémités soient achevées (Gratias and Betermier, 2003; Gratias et al., 2008) (Figure 15). Le mécanisme de NHEJ permettant une réparation très précise des cassures double brin, des phases de lecture fonctionnelles peuvent être restaurées après excision, sans introduire de mutations potentiellement délétères dans les gènes.

IV. Ciblage de la machinerie d'excision aux régions éliminées

Le consensus faiblement conservé présent aux extrémités des IES n'est pas suffisant pour expliquer le ciblage de l'endonucléase PiggyMac au niveau d'un tel nombre de séquences à éliminer. Le consensus pourrait en revanche favoriser la précision et l'efficacité des cassures. Comment, alors, l'élimination des séquences micronucléaires peut-elle être conduite de façon si efficace à chaque cycle de reproduction ?

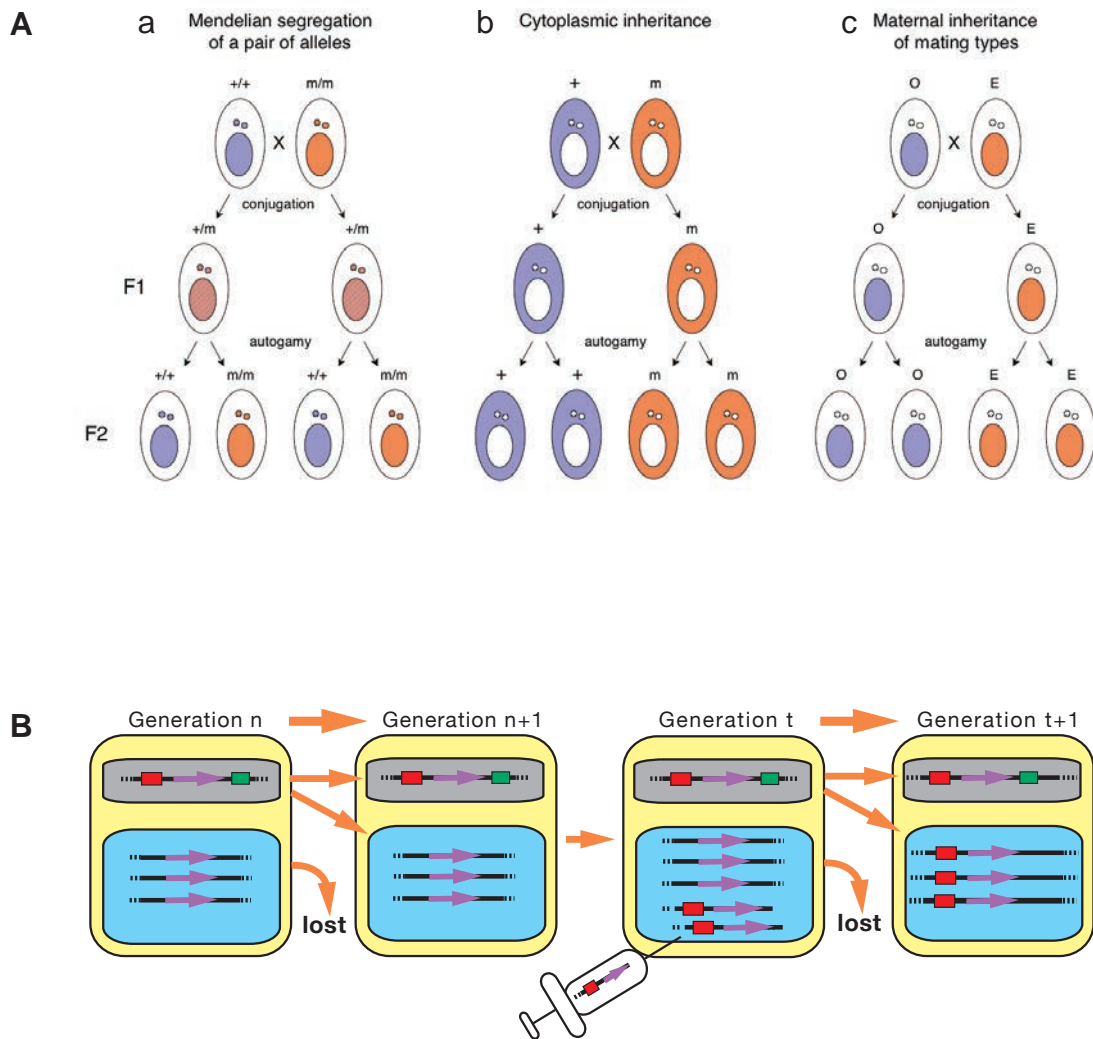


Figure 16. A. Hérité mendélienne et non mendélienne. **a.** Ségrégation mendélienne de deux paires d'allèles. La conjugaison entre deux cellules homozygotes pour différentes paires d'allèles à un locus (+/+ et m/m) génère des individus hétérozygotes (+/m) identiques en F1. L'autogamie de ces clones de F1 conduit à des individus complètement homozygotes en F2 qui ont 50% de chance de retenir chacun des allèles parentaux. **b.** Hérité cytoplasmique de caractères (par exemple, caractères déterminés par le génome mitochondrial). Très peu de cytoplasme est échangé entre les deux partenaires pendant la conjugaison. **c.** Hérité maternelle du type sexuel. L'hérité du type sexuel suit celle des traits cytoplasmique plutôt qu'une hérité mendélienne, bien que le type sexuel soit déterminé par le génome macronucléaire mature. **B. Inhibition homologie-dépendante de l'excision des IES par le macronoyau maternel.** Pendant le développement normal, les IES (rectangles rouge et vert) sont éliminées efficacement du génome macronucléaire. La transformation du macronoyau maternel avec une IES donnée peut inhiber l'élimination de l'IES homologue dans les nouveaux macronoyaux en développement.

Beisson et al, Cold Spring Harb Protoc 2010

Meyer Chalker, 2006 Epigenetics of ciliates

1. Régulation épigénétique des réarrangements du génome

De nombreuses études génétiques conduites chez *Paramecium tetraurelia* et chez *Tetrahymena thermophila*, un cilié apparenté à *Paramecium*, ont montré que le génome somatique maternel pouvait déterminer l'hérédité de caractères indépendamment du génotype germinale. Au fil du temps, il a été reconnu que cet effet épigénétique était médié par la régulation d'événements de réarrangements du génome. Le premier cas de ce type à avoir été compris au niveau moléculaire est le cas de l'épimutante d48 chez *Paramecium* (Epstein and Forney, 1984). Cette lignée de cellules est incapable d'exprimer l'antigène de surface A en raison d'une délétion à l'extrémité d'un chromosome macronucléaire qui est héritée d'une génération à l'autre, bien que la copie germinale du gène soit parfaitement sauvage (Forney and Blackburn, 1988). Ce caractère est donc soumis à une hérédité cytoplasmique, plutôt que mendélienne.

Les deux types sexuels de *P. tetraurelia*, nommés O et E sont également soumis à une hérédité cytoplasmique. Après la conjugaison, le type sexuel d'un clone végétatif descendant d'un parent O est presque toujours O et le type sexuel d'un clone descendant d'un parent E est presque toujours E alors que les deux cellules ex-conjugantes se sont développées à partir du même génome zygotique (Figure 16A). De plus, lorsqu'un large pont cytoplasmique se forme entre les deux cellules conjugantes, permettant un échange significatif de cytoplasme, la progéniture des deux parents est généralement de type E (Sonneborn, 1977). Il existe donc un facteur cytoplasmique qui dirige le développement du type sexuel E.

L'excision des IES est aussi soumise à un contrôle maternel. Chez *Paramecium* comme chez *Tetrahymena*, l'introduction d'une séquence d'IES dans le macronoyau maternel conduit à la rétention de la séquence germinale homologue dans le macronoyau de la descendance sexuelle, tandis que l'excision des autres IES n'est pas affectée (Chalker and Yao, 1996; Duhaucourt et al., 1995) (Figure 16B). Chez *Paramecium*, ce contrôle maternel de l'inhibition de l'excision a été observé pour 1/3 des IES testées qui ont été appelées en conséquence IES maternellement contrôlées (mcIES). Toutefois, l'effet maternel varie quantitativement parmi les IES. La fréquence d'élimination dépend de la dose injectée, celle-ci augmentant avec le nombre de copies germinales injectées (Duhaucourt et al., 1998). Dans certains cas, la rétention d'une IES résulte en une lignée cellulaire différente dans laquelle l'IES est maintenue stablement dans le génome macronucléaire d'une génération sexuée à l'autre, bien que le génome germinale soit complètement sauvage (lignée IES+).

Des profils d'excision alternatifs peuvent donc être hérités épigénétiquement en fonction du contenu du macronoyau maternel. Cela suggère que les éléments à éliminer ne sont pas définis par leur séquence d'ADN ou leurs séquences adjacentes mais par leur absence du génome somatique maternel. Ces observations ont conduit à proposer que la reconnaissance des éléments éliminés soit basée sur une comparaison en *trans* entre le génome germlinal à réarranger et le génome du macronoyau maternel déjà réarrangé, qui est toujours présent dans le cytoplasme des cellules pendant le développement des nouveaux macronoyaux. En conséquence, n'importe quelle séquence ne se trouvant pas dans le macronoyau maternel sera ciblée pour l'élimination. Ce mécanisme de comparaison présentant une forte spécificité de séquence, il pourrait être médié par des hybridations entre acides nucléiques de séquences homologues (Meyer and Garnier, 2002).

2. ARN non codants et réarrangements programmés du génome

a. ARN germinaux maternels : les scanARN

Une avancée majeure dans la compréhension des mécanismes de régulation épigénétique des réarrangements est venue de la découverte d'une voie de l'ARN interférence spécifique des événements sexuels et requise pour les réarrangements du génome (Mochizuki et al., 2002). Chez *Paramecium*, des petits ARN non codants de 25 nucléotides de long, les scanARN, sont produits à partir du génome micronucléaire au moment de la méiose des micronoyaux par une voie de l'ARN interférence spécifique de la méiose (Lepère et al., 2009). La biogenèse des scanARN requiert des protéines apparentées aux protéines Dicer et possédant une activité RNase-III : Dicer-like 2 (Dcl2) et Dicer-like 3 (Dcl3). Le séquençage des scanARN a révélé qu'une majorité de ces petits ARN présentait une signature 5'-UNG tandis qu'une fraction mineure présente la signature complémentaire aux positions 21-23. Les scanARN sont donc clivés à partir de précurseurs double brin et les ARN simple brin 5'-UNG sont ensuite préférentiellement stabilisés (Lepère et al., 2009).

La population de scanARN est très complexe et ne contient pas exclusivement des séquences spécifiques du génome micronucléaire. En effet, les scanARN sont produits à partir d'une large fraction du génome germlinal incluant exons, introns, régions intergéniques, transposons et IES. Les données obtenues jusqu'à présent sont cohérentes avec l'idée que le génome micronucléaire

est entièrement transcrit pour produire les scanARN. L'inactivation conjointe de Dcl2 et Dcl3 conduit à la disparition des scanARN, des défauts d'élimination d'ADN et à la mort de la progéniture sexuelle des cellules (Lepère et al., 2009). L'évidence la plus directe que les scanARN peuvent cibler l'élimination des IES vient d'expériences de micro-injection d'ARN synthétiques de 25 nucléotides qui miment la structure des scanARN et provoquent l'élimination de leurs séquences homologues dans le macronoyau en développement (Lepère et al., 2008).

Les scanARN sont associés à des protéines Piwi, membres de la famille Argonaute. Chez *Paramecium*, deux protéines Piwi, Ptiwi01 et Ptiwi09 sont requises pour l'accumulation des scanARN, pour l'élimination des séquences spécifiques du génome micronucléaire et la survie de la progéniture sexuelle des cellules (Bouhouché et al., 2011). Chez *Tetrahymena*, les scanARN produits par la protéine Dicer-like 1 (Dcl1p) s'associent à la protéine Piwi Twi1p (Malone et al., 2005; Mochizuki et al., 2002).

Il existe de fortes similitudes entre les scanARN et les piARN des métazoaires. Structuralement, ces deux types d'ARN non codant sont plus longs que les siARN et ont un biais de U en 5'. Chez *Tetrahymena*, les scanARN sont, comme les piARN, méthylés par l'ARN méthyltransférase Hen1. Fonctionnellement, scanARN et piARN sont tous deux requis pour la reproduction sexuée et pour le contrôle d'éléments transposables (ou leurs restes présumés pour les IES). Les piARN le font par l'intermédiaire d'une répression épigénétique et les scanARN par l'intermédiaire d'une élimination épigénétique. Pourtant, il existe des différences entre les deux types de mécanismes. Les scanARN agissent après la formation du noyau zygotique, pendant la formation du génome somatique et ne sont pas connus pour être requis pour la répression du génome germinale, contrairement aux piARN. A l'inverse des scanARN, les protéines Dicer-like ne semblent pas requises pour la biogenèse des piARN (Li and Liu, 2011). De plus, les scanARN ne montrent aucune évidence d'un mécanisme d'amplification par « ping-pong » impliquant l'activité endoribonucléolytique des protéines Piwi, comme c'est le cas pour les piARN (Brennecke et al., 2007; Gunawardane et al., 2007).

b. ARN non codants somatiques maternels

Si les scanARN sont produits à partir de la totalité du génome micronucléaire, comment peuvent-ils cibler uniquement l'élimination des séquences absentes du génome somatique maternel ?

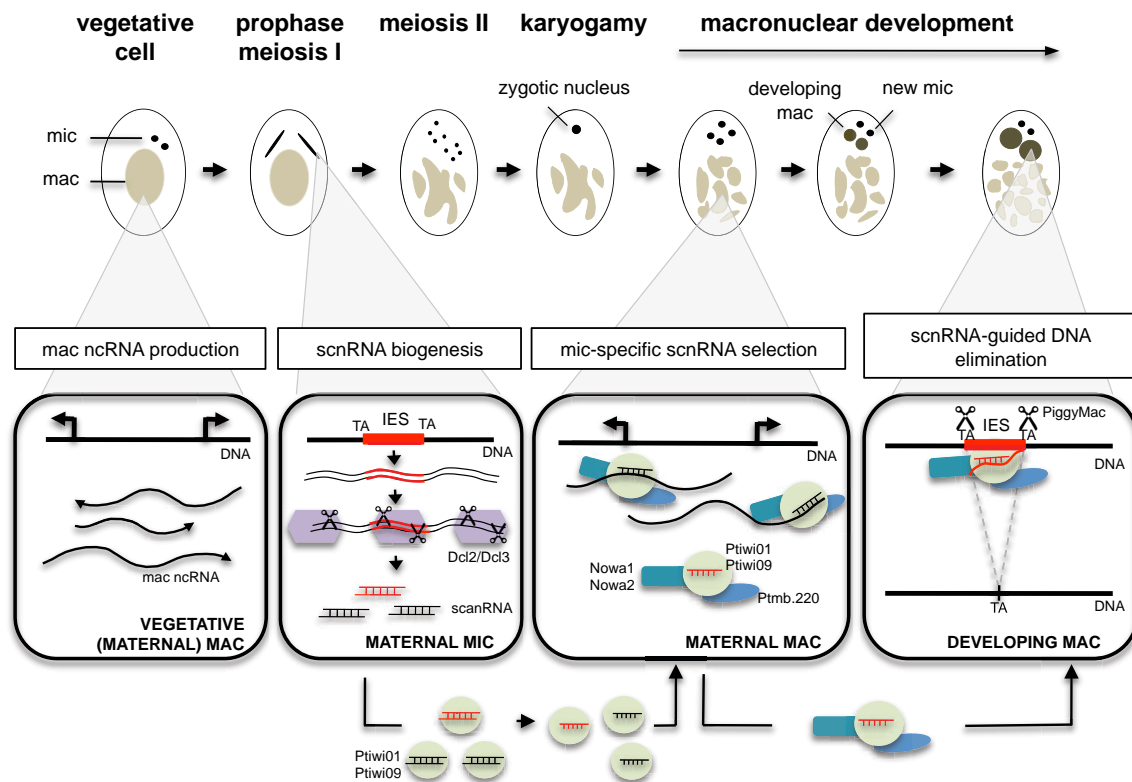


Figure 17. Le modèle de « genome scanning » chez *Paramecium*. La progression des événements de développement nucléaires est représenté en haut, les étapes correspondantes du processus de scanning et les facteurs impliqués sont montrés en dessous. 1. Transcription maternelle non codante. Des transcrits non codants sont produits constitutivement à partir des deux brins d'ADN du génome macronucléaire entier. 2. Biogenèse des scanARN dans les micronoyaux en méiose. Les duplexes de scanARN produits par Dcl2 et Dcl3 sont transportés dans le cytoplasme où ils sont chargés dans les complexes Ptiwi01 et Ptiwi09 conduisant à l'élimination du brin « passager ». 3. Sélection des scanARN mic-spécifiques. Les complexes Ptiwi-scanARN sont transportés dans le macronoyau maternel. L'appariement par homologie de séquences entre les scanARN et les transcrits macronucléaires non codants conduit à la sélection des scanARN portant des séquences spécifiques qui seront transportées dans le macronoyau en développement. Le processus de scanning pourrait être favorisé par l'hélicase Ptmb.220 et assisté par les protéines de fixation à l'ARN Nowa1 et Nowa2. 4. Elimination d'ADN. Les scanARN reconnaissent leurs séquences homologues dans le génome du macronoyau en développement ciblant le recrutement de l'endonucléase PiggyMac.
Coyne et al., *Biol Cell*, 2012

En opposition avec les scanARN qui semblent promouvoir l'élimination de leurs séquences d'ADN homologues, une autre classe de transcrits non codants semble avoir un rôle protecteur contre l'élimination. Des expériences de RT-PCR ont montré que des transcrits non codants sens et antisens sont produits constitutivement, en faible quantité, à partir du génome macronucléaire. La dégradation ciblée de ces transcrits par ARN interférence conduit à la délétion des séquences d'ADN homologues dans le macronoyau en développement (Lepère et al., 2008). De manière similaire, l'inactivation par ARN interférence d'un gène non essentiel pendant la croissance végétative déclenche la délétion de ce gène dans le macronoyau en développement après la reproduction sexuée (Garnier et al., 2004). Le ciblage d'une petite région du gène suffit à déclencher la délétion du gène entier, suggérant que l'inactivation par ARN interférence conduit à la dégradation des transcrits couvrants le gène complet. Par conséquent, pour éviter d'être éliminées dans le macronoyau en développement, les séquences doivent non seulement être présentes dans le génome maternel mais aussi générer des transcrits qui vont contrecarrer l'action des scanARN. A première vue, cette situation est analogue à la situation observée chez la levure *S. pombe*, les plantes et les animaux où la transcription active de certaines régions génomiques est nécessaire pour la formation d'hétérochromatine et la répression de ces séquences (Castel and Martienssen, 2013).

3. Le modèle de « genome scanning »

Pour expliquer comment ces deux classes d'ARN non codants maternels aux fonctions antagonistes agissent et régulent épigénétiquement l'élimination des séquences d'ADN mic-spécifiques, un modèle de « genome scanning » a été proposé. Selon ce modèle, les scanARN scruteraient le génome somatique en s'appariant avec les transcrits non codants somatiques. Ce processus permettrait la sélection des scanARN portant des séquences mic-spécifiques, en d'autres termes ceux qui n'ont pas réussi à trouver de séquence homologue dans le génome somatique maternel. Ces derniers pourraient alors guider la délétion des séquences germinales dans le macronoyau en développement. L'appariement entre les scanARN, promouvant la délétion et les transcrits somatiques, protégeant de la délétion constituerait une soustraction génomique naturelle permettant l'élimination de toutes les séquences absentes du macronoyau maternel (Figure 17) (Coyne et al., 2012).

De nombreuses études conduites chez *Paramecium* et *Tetrahymena* ont permis de soutenir cette hypothèse et d'en apercevoir les mécanismes sous-jacents.

Les protéines associées aux scanARN Ptiwi09 (chez *Paramecium*) et Twilp (chez *Tetrahymena*) sont localisées dans le macronoyau maternel avant que les nouveaux macronoyaux ne se développent, soutenant l'idée que le processus de scanning se déroule dans le macronoyau maternel (Bouhouche et al., 2011; Mochizuki et al., 2002). Après leur production dans les micronoyaux en méiose, les duplexes de scanARN seraient transportés dans le cytoplasme et chargés dans les protéines Piwi. Chez *Tetrahymena*, il a été montré que Twilp possède, comme la plupart des protéines Argonaute, une activité « slicer » qui permet de cliver un brin du duplex de scanARN. Cette activité permet l'éviction du brin dit « passager » dans les complexes Twilp-scanARN. L'activité « slicer » de Twilp est aussi requise pour son import dans le macronoyau maternel. Twilp complexé avec les scanARN simple brin serait transporté dans le macronoyau parental avec l'aide d'une autre protéine d'interaction, Gwilp (Noto et al., 2010). Ces scanARN sont ensuite 2'-O-méthylés par Hen1p, une ARN méthyltransférase localisée dans le macronoyau maternel (Kurth and Mochizuki, 2009).

Dans le macronoyau maternel, l'appariement des scanARN et des transcrits non codants maternels permettrait de sélectionner les scanRNA portant des séquences mic-spécifiques à partir de la population initiale très complexe de scanARN. Les scanARN qui correspondent à des séquences macronucléaires seraient hybridés avec les transcrits maternels complémentaires et cette interaction conduirait d'une certaine manière à leur inactivation. A l'inverse, les scanARN qui correspondent à des séquences spécifiques du génome micronucléaire ne trouveraient par définition pas de séquence homologue parmi les transcrits somatiques maternels. Ils seraient donc libres d'être transportés dans le noyau zygotique, dès que celui-ci se développe, pour aller cibler l'élimination de leurs séquences homologues. Les mécanismes par lesquels les scanARN seraient inactivés ne sont pas connus. Une possibilité serait que les transcrits maternels non codants agissent comme une éponge moléculaire en séquestrant les scanRNA auxquels ils sont appariés. L'interaction entre les deux types de transcrits pourrait également conduire à la dégradation des scanARN. En accord avec cette dernière hypothèse, la population de scanARN semble s'enrichir en séquences mic-spécifiques au cours des événements sexuels (Lee and Collins, 2006; Lepère et al., 2009; Mochizuki and Gorovsky, 2004; Sandoval et al., 2014).

Un certain nombre de protéines essentielles au processus de scanning ont été identifiées. Chez *Tetrahymena*, l'hélicase putative Ema1p est essentielle pour l'interaction entre les scanARN, Twi1p et les transcrits maternels macronucléaires naissants et semble permettre la dégradation de scanARN portant des séquences macronucléaires (Aronica et al., 2008). La dissociation de Twi1p des scanARN appariés avec des transcrits somatiques pourrait conduire à la dégradation des scanARN et au recyclage de Twi1p (Aronica et al., 2008). Chez *Paramecium*, l'hélicase putative PtmB.220p est requise pour les réarrangements du génome et pourrait jouer le même rôle que son homologue Ema1p (Nowak et al., 2011). Les protéines Nowa1p et Nowa2p, qui contiennent des répétitions GW caractéristiques de protéines interagissant avec des ARN, sont requises pour l'élimination des mclES et des transposons. Elles pourraient participer au processus de « scanning » mais des études supplémentaires sont nécessaires pour préciser leur fonction (Nowacki et al., 2005).

Un certain nombre de facteurs présentés précédemment incluant Nowa1p, Twi1p, Ptiwi09p et Ema1p sont localisés dans le macronoyau maternel avant l'achèvement de la méiose des micronoyaux et sont ensuite transportés dans les macronoyaux en développement. Cela suggère que ces protéines sont impliquées dans le transport des scanARN qui auraient été sélectionnés vers les macronoyaux en développement. Ces derniers pourraient alors guider l'élimination de leurs séquences homologues.

4. Ciblage de l'endonucléase par les scanARN au niveau des séquences à éliminer

Chez *Tetrahymena*, deux marques spécifiques de la chromatine répressive, H3K9me3 et H3K27me3 sont associées aux IES pendant les événements de réarrangements du génome (Liu et al., 2007; Taverna et al., 2002). Ces modifications sont requises pour leur élimination puisque des mutations dans la séquence de H3 empêchant la méthylation de K9 ou de K27 conduisent à la rétention des IES (Liu et al., 2004, 2007). Une lysine méthyltransférase à domaine SET, Ezl1p, homologue de Enhancer of zeste chez la *Drosophile* régule la déposition de ces marques pendant le développement et est aussi requise pour l'élimination des IES (Liu et al., 2007). La méthylation par Ezl1 est dépendante de la voie des scanARN. En effet, la déplétion de Twi1p abolit la méthylation de H3K27 et de H3K9 dans le macronoyau en développement (Liu et al., 2004, 2007; Malone et al., 2005). Deux protéines à chromodomaine spécifiques du cycle sexuel, Pdd1

et Pdd3, fixent ces marques d'histones (Pdd1 reconnaît H3K9me3 et H3K27me3 tandis que Pdd3 reconnaît seulement H3K9me3) déclenchant la formation de structures hétérochromatiniennes à la périphérie du noyau. Ces structures sont essentielles pour l'élimination des IES et permettraient le recrutement de l'endonucléase homologue de PiggyMac, Tpb2 (Cheng et al., 2010). Les marques d'histones pourraient également recruter Tpb2 directement par l'intermédiaire de son domaine PHD (Vogt and Mochizuki, 2013).

Ce mécanisme d'élimination de l'ADN dirigé par les scanARN présente de fortes similarités avec la formation d'hétérochromatine chez les autres eucaryotes, excepté qu'il implique une lysine méthyltransférase de type Polycomb et non une protéine de la famille SU(VAR)3-9. Il va également une étape plus loin, avec l'élimination physique des séquences qui ont été hétérochromatinisées.

Chez *Paramecium*, les facteurs qui pourraient être impliqués dans le recrutement de l'endonucléase par les scanARN dans les macronoyaux en développement n'ont pas encore été décrits. Une autre classe d'ARN non codants, les iesARN, a cependant été identifiée récemment et pourrait améliorer l'efficacité du processus d'élimination dans le macronoyau en développement. La biogenèse des iesARN dépend de la protéine Dicer-like 5 (Dcl5p) et la séquence de ces petits ARN de 26-30 nucléotides correspond exclusivement aux IES. La déplétion de Dcl5p conduit à la rétention d'un sous-ensemble d'IES. Il a été proposé que la transcription des IES excisées serait à l'origine de la production des iesARN par Dcl5p. Les iesARN permettrait de promouvoir l'excision des IES encore maintenues dans le génome en établissant une boucle de rétroaction positive entre production d'iesARN et élimination d'ADN (Sandoval et al., 2014).

OBJECTIF DE THESE N°1 :

Déterminer le rôle des modifications d'histones dans le recrutement de l'endonucléase PiggyMac au niveau des séquences à éliminer chez *Paramecium tetraurelia*

Chez *Paramecium*, le jeu de séquences à éliminer est plus complexe en taille et en séquence que chez *Tetrahymena*. En effet, le génome micronucléaire de *Tetrahymena* ne contient qu'environ 6000 IES qui sont majoritairement longues (leur taille varie entre 1 et 2 kb), qui peuvent être répétées et dont l'excision n'est pas précise. Les 45000 IES de *Paramecium* sont à l'inverse uniques, courtes et requièrent une excision précise. Pour des IES aussi courtes que 26 pb et donc plus courtes que la taille de l'ADN qui entoure un nucléosome, comment peut-on envisager le rôle de modifications d'histones pour définir ces séquences ? De manière générale, la méthylation des histones est-elle requise pour l'élimination des séquences mic-spécifiques chez *Paramecium* ?

Dans la revue (Coyne et al., 2012), nous avons fait l'hypothèse que les séquences les plus longues (au moins supérieures à 150 pb) pourraient être reconnues grâce à des modifications d'histones tandis que les séquences les plus courtes (inférieures à 150 pb) seraient définies par d'autres mécanismes comme des modifications de nucléotides, ou par les complexes Ptiwi-scanARN directement.

Pour tester l'hypothèse selon laquelle la méthylation des histones interviendrait dans l'élimination des séquences les plus longues, nous avons cherché la présence de marques d'histones répressives pendant le développement macronucléaire de *Paramecium* et cherché la présence de lysine méthyltransférases spécifiques du développement. Nous avons identifié une protéine de la famille Enhancer of zeste, homologue de Ezh1 chez *Tetrahymena*, requise pour l'élimination d'un sous-ensemble de séquences d'ADN pendant le développement. Des analyses à l'échelle du génome entier montrent qu'il existe en effet un biais de taille dans les séquences qui dépendent de EZH1 pour être éliminées (Voir Résultats, article 1, Lhuillier-Akakpo et al., en préparation).

V. Dimorphisme nucléaire et division cellulaire

Comme nous venons de le voir, micronoyaux et macronoyau se distinguent par leur contenu génétique et le développement du génome macronucléaire est associé à d'importants réarrangements génomiques. De manière intéressante, il a été observé depuis longtemps, par des expériences de microscopie électronique, que les deux types de noyaux diffèrent également par leur type de ségrégation. En effet, pendant la croissance végétative des cellules, les micronoyaux ségrègent par mitoses classiques tandis que le macronoyau se divise par un processus d'élongation non mitotique qui n'impliquerait ni condensation des chromosomes ni formation de fuseau mitotique (Tucker et al., 1980). Les chromosomes macronucléaires seraient répartis aléatoirement entre les cellules filles et la forte ploïdie du génome macronucléaire pourrait permettre de maintenir un nombre suffisant de chaque chromosome à chaque division cellulaire. Les chromosomes macronucléaires se comportent donc comme s'ils étaient dépourvus de centromères actifs.

Peu de choses sont connues au sujet des centromères chez les ciliés. Aucune séquence centromérique n'a jamais été identifiée parmi les organismes de ce phylum. Chez *Tetrahymena*, le variant d'histone centromérique, Cna1p, a été identifié et se trouve localisé uniquement dans les micronoyaux. Il est essentiel aux fonctions micronucléaires indiquant que seuls les chromosomes micronucléaires contiennent des centromères actifs (Cervantes et al., 2006; Cui and Gorovsky, 2006). La protéine de fusion Cna1p-GFP disparaît des centromères dans les macronoyaux en développement et il a été suggéré que les séquences centromériques fassent partie des séquences micronucléaires éliminées (Cui and Gorovsky, 2006).

OBJECTIF DE THESE N°2 :

Comprendre les mécanismes d'inactivation programmée des centromères au cours du développement macronucléaire de *Paramecium tetraurelia*

Pour déterminer comment est régulée l'activité centromérique dans les noyaux des cellules de *Paramecium tetraurelia*, nous avons cherché à identifier le variant d'histone centromérique CenH3 chez cet organisme.

Nous avons identifié deux gènes codants pour des protéines CenH3 putatives dont une, CenH3-1 se localise uniquement dans les micronoyaux et est essentielle aux fonctions micronucléaires. En utilisant CenH3-1 comme marqueur des centromères actifs, nous avons déterminé que l'inactivation des centromères germinaux se produisait pendant le développement macronucléaire, au moment où ont lieu les réarrangements du génome.

Deux hypothèses sont donc possibles pour expliquer le phénomène d'inactivation des centromères:

1. Les centromères germinaux sont inactivés épigénétiquement, par un mécanisme qui conduit à l'éviction de CenH3 des chromosomes macronucléaires pendant le développement.
2. Les centromères germinaux sont inactivés par l'élimination physique des séquences qui les portent, par des mécanismes associés aux réarrangements du génome.

Nous avons notamment testé cette deuxième hypothèse en testant l'effet de gènes impliqués dans l'élimination programmée d'ADN sur le processus d'inactivation des centromères (Voir Résultats, article 2, en préparation). Pour valider l'une des deux hypothèses, nous avons également tenté d'identifier les séquences centromériques de *Paramecium* en identifiant par ChIP les séquences associées au variant d'histone CenH3.

RESULTATS

PARTIE 1

Rôle de la méthylation des histones dans les réarrangements programmés du génome chez *Paramecium tetraurelia*

ARTICLE 1 (en préparation)

Chez *Paramecium tetraurelia*, la différenciation du génome somatique s'accompagne de l'élimination physique de nombreuses séquences d'ADN germinales. Le ciblage de l'endonucléase PiggyMac au niveau des séquences à éliminer fait intervenir une voie de l'ARN interférence spécifique du développement mais les mécanismes impliqués ne sont pas bien compris.

Dans l'article qui suit, nous mettons en évidence que la triméthylation de H3K27 et de H3K9, deux marques associées à la chromatine répressive, sont présentes transitoirement dans les macronoyaux en développement de *P. tetraurelia*. Nous avons identifié une lysine méthyltransférase, apparentée aux protéines Enhancer of zeste (E(z)) et nommée Ezl1p, requise pour la biogenèse de ces modifications d'histone. Par des analyses à l'échelle du génome entier, nous avons montré que la déplétion de Ezl1p conduisait à la rétention de toutes les séquences micronucléaires imprécisément éliminées et empêchait l'excision précise d'un sous-ensemble d'IES, suggérant que les marques d'histones sont requises pour l'action de l'endonucléase PiggyMac. De manière frappante, il existe un biais de taille entre les IES sensibles et les IES insensibles à EZL1, les IES les plus longues étant les plus affectées. La voie des scanARN est requise pour la déposition de H3K27me3 et de H3K9me3 dans les macronoyaux en développement mais la déplétion des scanARN n'affecte également qu'un sous-ensemble d'IES.

L'ensemble de nos données suggère que les modifications d'histones H3K27me3 et H3K9me3 participent au recrutement de l'endonucléase PiggyMac au niveau des régions à éliminer. Les différences de sensibilités des IES face à l'inactivation de EZL1 et face à la déplétion des scanARN suggèrent qu'il existe des voies partiellement chevauchantes pour le recrutement de PiggyMac.

**The histone methyltransferase EZL1 is required for
programmed genome rearrangements and displays a DNA length bias**

Maoussi Lhuillier-Akakpo^{1,2,*}, Andrea Frapporti^{1,*}, Cyril Denby Wilkes³, Mélody Matelot¹,
Michel Vervoort¹, Linda Sperling³ & Sandra Duhaucourt¹

¹Institut Jacques Monod, CNRS, UMR 7592, Université Paris Diderot, Sorbonne Paris Cité, Paris, F-75005 France.

²Sorbonne Universités, UPMC Univ., IFD, 4 place Jussieu, 75252 Paris cedex05, France.

³CNRS UPR3404 Centre de Génétique Moléculaire, Gif-sur-Yvette F-91198 cedex, France; Université Paris-Sud, Département de Biologie, Orsay, F-91405, France.

*These authors contributed equally to this work

Correspondence should be addressed to S. D. (duhaucourt@ijm.univ-paris-diderot.fr)

Key words: genome rearrangements, histone methylation, sRNAs, ciliates, *Paramecium*

RESULTS

Dynamic localization of histone H3 tri-methylation on K9 and K27 in the developing MAC

Indirect immunostaining experiments using commercial anti- H3K27me3 and -H3K9me3 antibodies were performed to determine the *in situ* localization of both marks during various stages of *Paramecium* life cycle. Specificity of the antibodies was verified by dot blot and competition assays (Fig. S1). No H3K27me3 or H3K9me3 could be detected in the transcriptionally active MAC or in the transcriptionally inactive MIC during vegetative growth (Fig. S2). The sexual process of autogamy (self-fertilization), which is induced by starvation, starts with meiosis of the micronuclei and proceeds through the development of new MACs (Fig. 1 and S2). Karyogamy gives rise to a diploid zygotic nucleus that divides twice; two of the products differentiate into new micronuclei and the other two into new macronuclei. The maternal macronucleus becomes fragmented shortly after micronuclear meiosis. Because cells enter autogamy from a fixed point of the cell cycle (Berger, 1986), a minimum asynchrony of 5–6 h is observed between the first and the last cells to undergo meiosis. Progression through autogamy was monitored by counting cells at different stages after Hoechst staining (Fig. S3). During sexual events, H3K27me3 was transiently found in the MIC during the first meiotic division and detected in the fragments of the maternal MAC by the end of meiosis, whereas no H3K9me3 signal was observed at these stages (Fig. S2). Both H3 K9me3 and K27me3 were detected at early stages of MAC development and the signal persisted throughout the course of MAC development, when genome rearrangements occur (Fig. 1 and S2). The enrichment of H3K27me3 in the developing MAC was also confirmed by Western blot analysis on purified nuclei (Fig. S2).

The staining in the developing MAC is very similar with both anti-H3K9me3 and anti-H3K27me3 antibodies. Initially diffuse and evenly distributed (Fig. 1 panel a) the staining gradually condensed into a punctuate pattern (Fig. 1 panels b, c, d). The H3K9me3 and H3K27me3 foci resemble the heterochromatin bodies detected in the distantly related ciliate *Tetrahymena*, which comprise H3K9me3, H3K27me3 (Liu et al., 2004, 2007) and the chromodomain protein Pdd1p (Madireddi et al., 1996). Yet these foci are located inside the nucleus (not shown) and are thus not preferentially located at the periphery of the developing macronucleus as observed in *Tetrahymena*. As development proceeds, the number of these intensely labeled foci is reduced and the single remaining spot that

is found in a DNA-poor region of the macronucleus (Fig. 1d and S2), eventually disappeared.

The domesticated transposase PiggyMac is required for H3K9me3 and H3K27me3 foci formation

To obtain further insight into the possible role of H3K9me3 and H3K27me3 in genome rearrangements, we knocked-down by RNA interference (RNAi) the domesticated transposase PiggyMac (Pgm) that is required for the introduction of DNA-double strand breaks at the boundaries of germline-limited segments (Arnaiz et al., 2012; Baudry et al., 2009). RNAi can be obtained by feeding *Paramecium* cells on *E. coli* bacteria that have been engineered to produce double-stranded RNA from the target gene (Galvani and Sperling, 2002). As expected, Pgm depletion led to retention of all tested germline-limited elements in the developing MAC genome and to 100% lethality in the sexual progeny (not shown). Immunofluorescence experiments revealed that H3K9me3 and H3K27me3 signals progressively increased in the developing MAC and are detected in all cells in Pgm-depleted and control cells (Fig. 2A and S3). Western blot analysis showed that the amount of H3K27me3 is not significantly altered in Pgm-depleted cells compared to control cells (Fig. 2B). Because depletion of Pgm does not affect the biogenesis and accumulation of H3K9me3 and H3K27me3 in the developing MAC, we conclude that the endonuclease Pgm acts downstream of H3K9me3 and H3K27me3, in agreement with the scanning model.

The initially diffuse H3 K9me3 and K27me3 signals became punctuate in control cells. But both H3 K9me3 and K27me3 signals remained diffuse as development proceeds in Pgm-depleted cells: no foci could be detected at any time points (Fig. 2A). We conclude that the endonuclease Pgm is required for H3 K9me3 and K27me3 foci formation. At this stage, it is not clear whether the foci are the result of DNA double strand break formation or are a prerequisite for DNA elimination to occur.

The scnRNA pathway is required for H3K27me3 and H3K9me3 accumulation in the early developing MAC

The scanning model posits that MIC-specific scnRNAs guide the loading of histone marks specifically on DNA segments that are eliminated in the developing MAC. We expected that the co-silencing of the two Dicer-like genes, DCL2 and DCL3, that results in failure to generate scnRNAs (Lepere et al., 2009; Sandoval et al., 2014), would also

abolish the establishment of H3 K9me3 and K27me3 chromatin in the developing MAC. To test this hypothesis, we examined the effects of DCL2/3 co-silencing on H3 K27me3 and K9me3 by immunofluorescence staining. As expected, the depletion of Dcl2 and Dcl3 proteins led to 90-100% lethality in the sexual progeny (not shown). No detectable H3 K9me3 or K27me3 signal was observed in Dcl2/3p depleted cells at an early stage when the developing MAC of control cells stained intensely (Fig. 2C). As development proceeds, increasing numbers of developing MAC with H3 K27me3 and K9me3 signals were observed in Dcl2/3p-depleted cells (Fig. S3), but the intensity of the immunofluorescence signal was much lower than in the control cells (not shown). Western blot analysis confirmed that the total amount of H3K27me3 is greatly reduced in DCL2/3-knockdowns (KD) relative to control (Fig. 2D). Therefore, we conclude that the generation of scnRNAs is required for the establishment and accumulation of these chromatin modifications in the developing MAC. These results suggest that scnRNAs and K9 and K27 methylation participate in the same pathway leading to genome rearrangements. However, to prove this suggestion, it is necessary to demonstrate that K9 and K27 methylation is required for DNA elimination.

Identification of SET-domain containing proteins and functional analysis of 5 EZL genes reveal that EZL1 is essential

To eliminate K9 and K27 methylation, we sought to identify the gene(s) responsible for these modifications. We searched for SET domain containing proteins encoded in the *P. tetraurelia* MAC genome and found 34 putative Lysine Methyl Transferases (KMTs) (Fig. S4, Table S1). No H3K9-specific KMTs homologs could be identified in the *P. tetraurelia*, *Tetrahymena thermophila* or *Oxytricha trifallax* genomes. We identified five putative H3K27-specific KMT, named EZL1 to EZL4. EZL3a and EZL3b are gene duplicates from the last whole genome duplication (Aury et al., 2006). Alignment of the predicted EZL proteins revealed conservation of key residues implicated in binding the methyl donor, the target lysine and catalysis (Fig. S5).

The expression patterns of EZL genes during the life cycle were examined using microarray data (Arnaiz et al., 2010) and confirmed by RT-PCR analysis from exponentially growing vegetative cells and from 7 different time points after the onset of autogamy in mass cultures (Fig. S6). Little or no expression is observed during vegetative growth but the genes are specifically expressed during the sexual phase of the life cycle, although they show markedly different patterns. EZL2 and EZL4 are silent

during vegetative growth but EZL4 is specifically expressed after meiosis, whereas EZL2 becomes expressed at the onset of MAC development. EZL1 gene is turned on to high levels immediately upon meiosis, and this is true also, to a much lesser extent, for EZL3a and EZL3b. Expression of EZL1 gene is very transient, preceded that of PGM, and detection of IES excision products (Fig. S6). This expression pattern is similar to that seen for the Dicer-like genes DCL2 and DCL3 (Lepere et al., 2009)(Fig. S6).

To test the function of EZL genes, we knocked down their expression by RNAi. Consistent with their specific expression during the sexual phase, continuous silencing of these genes during ≥ 15 vegetative divisions, alone or in combinations of 2 for the more closely related paralogs EZL3a and EZL3b, did not result in any obvious phenotype (not shown). Applying dsRNA feeding for only 4 divisions, immediately followed by autogamy, we observed cytologically normal new MAC developed in all cases, as monitored by Hoechst staining of cells. However, after EZL1 KD, 97% of post-autogamous progeny were unable to resume normal vegetative growth when returned to standard growth medium, whereas no lethality was observed in any other silencing conditions (Table S2).

The transcription of EZL1 is induced during meiosis, largely before programmed genome rearrangements take place in developing new MACs. This pattern led us to consider the possibility that this protein may be involved in a meiotic function. We thus checked the progression of meiosis by Hoechst staining during autogamy of EZL1 KD cells and observed that meiotic divisions I and II occur normally, since there were cells with 4 then 8 haploid nuclei in the population (Fig. S3), with no arrest until new MACs differentiate from mitotic copies of the zygotic nucleus.

To control for possible off-target silencing artifacts, two non-overlapping fragments of EZL1 were used independently to induce RNAi, and similar results were obtained with both constructs (Table S2). For one construct, the efficiency of EZL1 KD was checked by semi-quantitative RT-PCR of total RNA extracted throughout autogamy from control and EZL1 KD cells (Fig. S6): a significant decrease of EZL1 mRNA accumulation was observed at early time points in EZL1 KD cells, without affecting the onset of induction of other EZL genes. Therefore, EZL1 gene expression is essential during development for the production of viable sexual progeny.

In an EZL1 KD, the transcription of the PGM gene and of all EZL genes is switched on normally during autogamy, indicating that these genes are not induced in response to EZL1 induction but more likely as part of a general transcription program during MAC development. In contrast to control cells, the levels of these mRNAs do not decrease at

later time-points in an EZL1 silencing experiment (Fig. S6), suggesting that the completion of MAC development is a signal for transcriptional switch-off.

The Ezl1 protein is required for H3K9me3 and H3K27me3 in the developing MAC

To gain further insight into the role of EZL1, we examined the subcellular localization of Ezl1p. A GFP fusion was constructed by inserting the GFP coding sequence into the EZL1 gene, downstream of the start codon. Expression of the fusion gene was under the control of the natural EZL1 up- and downstream sequences. After microinjection of the construct into the MAC, no fluorescence could be detected during vegetative growth of transformed clones (Fig. S7A). When autogamy was induced, GFP fluorescence first appeared transiently in the MIC during meiosis I and in the MAC before it became fragmented. When fragmentation of the maternal MAC was complete, GFP fluorescence started to decrease and progressively relocated to the new MACs as they developed, until all of the fusion protein was concentrated there (Fig. 3A and S7). The localization pattern of GFP-Ezl1 fusion is very similar to that observed for H3K27me3 (Fig. 1 and S2). We observed that the GFP-Ezl1 fusion colocalized with H3 K27me3 and K9me3 foci in the new developing MACs (Fig. S7B). Moreover, the GFP-Ezl1 fusion showed the same behavior as H3 K9me3 and K27me3 signals: in PGM KD cells, no GFP-Ezl1 foci could be detected, and in DCL2/3 KD cells, the fluorescence intensity of the GFP fusion was greatly diminished (Fig. S7C).

We therefore investigated the effects of EZL1 KD on H3 K27me3 and K9me3 by examining these modifications by immunofluorescence staining and Western blot analysis. Immunofluorescence staining with H3K27me3- and H3K9me3-specific antibodies showed little or no signal in Ezl1p-depleted cells, whereas H3 K27me3 and K9me3 increased as development proceeded and completely disappeared at the latest time point (Fig. 3B and S3) in control cells. For more accuracy, we quantified the fluorescence intensities throughout the volume of the developing new MACs in control and EZL1 KD cells (see Materials and Methods). The nuclear signal was normalized to the fluorescence intensity of the same volume in the cytoplasm to measure noise. The mean ratio of the signal/noise was then calculated in each new MAC for H3 K9me3 or for K27me3 at different developmental time points (Fig. 3C-D). The quantification data clearly showed that both modifications were abolished in the developing new MACs of EZL1 KD cells. Western blot analysis of nuclear extracts during development confirmed that H3K27me3 was greatly reduced in Ezl1p-depleted cells (Fig. 3E). Together these

data showed that EZL1 encodes a developmental specific KMT responsible for H3 K27me3 and K9me3 in the developing new MAC.

The Ezl1 protein is required for imprecise deletion of MIC-specific sequences

EZL1 KD led to phenotypes that indicate an essential function for EZL1p during MAC development: (1) no viable sexual progeny were isolated from Ezl1p-depleted cells, a phenotype described in KDs defective in DNA elimination (Baudry et al., 2009; Bouhouche et al., 2011; Lepere et al., 2009; Nowacki et al., 2005; Nowak et al., 2011); (2) no H3 K9me3 and K27me3 were detected in the developing new MACs.

Different assays were thus used to monitor genome rearrangements in autogamy time course experiments, where the EZL1 gene was knocked-down by RNAi. We first tested the role of the EZL1 gene in the imprecise DNA elimination mechanism that is responsible for the deletion of MIC transposable elements during MAC development. In control RNAi experiments, only the rearranged forms originating from both the maternal and new MACs could be detected by Southern blot hybridization for the germline region located downstream from the surface antigen G gene, which contains a Sardine transposon that is eliminated imprecisely during MAC development, leading to chromosome fragmentation (Arnaiz et al., 2012)(Fig. 4A). In contrast, after EZL1 KD, non-rearranged forms accumulated throughout autogamy in the new MAC, relative to the rearranged forms present in the fragments of the maternal MAC (Fig. 4A). EZL1 KD led thus to retention of the MIC sequences located downstream from the normal telomere addition site and impaired germline chromosome fragmentation.

We further tested the role of Ezl1p in the imprecise elimination mechanism that is responsible for maternally inherited deletions of non-essential cellular genes, which can be induced experimentally (Garnier et al., 2004). The variant cell line 51DeltaND7 carries a maternally inherited MAC deletion of the ND7 gene but retains the gene in its MIC. We therefore used this strain to monitor the effect of EZL1 KD on maternal inheritance of MAC deletions. Reversion of the ND7 MAC deletion in post-autogamous cell populations was assessed by phenotypic testing: in EZL1 silencing but neither in the control silencing nor in the un-silenced control, the occurrence of trichocyst discharge indicated that the ND7 gene was at least partially maintained in the new MAC (not shown). Molecular analysis showed that it was indeed the case. ND7 was transiently amplified before deletion from the new developing MACs and, at later time points, only the rearranged forms, originating from both the old and new MACs, could be detected in control silencing

(Fig. 4B). In contrast, full-length ND7 gene product accumulated at late time points in EZL1 KD. Thus, like imprecise deletion of MIC specific regions, maternally inherited elimination of the ND7 gene is blocked in EZL1 KD cells, and the non-rearranged germline locus is retained in the developing new MAC.

To confirm genome-wide the effects of Ezl1p depletion on the imprecise elimination of germline DNA, we sequenced DNA isolated from newly developed MACs following EZL1 silencing. DNA isolated from newly developed MACs at the same developmental stage from a cell culture that was not submitted to RNAi was also sequenced as a control sample (Table S3). Because there is no good assembly of the MIC genome yet available, we compared the sequence complexity of different datasets by mapping the reads in each dataset to contigs previously assembled from new MACs after Pgm depletion (Arnaiz et al., 2012), currently the best representation of the un-rearranged germline genome. As shown in Table S4, EZL1 reads have the same sequence complexity as the PGM reads, while the control dataset has about 13 Mb less sequence complexity. Since the analysis only used Pgm contigs larger than 1 kb, the total sequence complexity in the MIC is expected to be larger than the 89 Mb we were able to analyze in this way.

If we compare the complexity for the regions that are not covered by the control sample, and correspond to the part of the MIC genome that is not collinear with MAC chromosomes, and most likely correspond to MIC specific DNA, PGM and EZL1 datasets show again a similar complexity. We also performed a qualitative evaluation of Sardine retention by mapping reads from each dataset to the known cloned copies of this transposable element (Arnaiz et al., 2012)(Fig. S8). Consistent with our Southern blot analysis (Fig. 4A), we found that all Sardine copies are significantly retained after EZL1 silencing. Therefore this global analysis supports the conclusions made at the molecular level for two individual loci: EZL1, like PGM, is required for the imprecise elimination of germline-limited sequences.

The Ezl1 protein is required for the excision of a subset of IESs

We then investigated the role of the EZL1 gene in IES excision. Excision was first analyzed by PCR on genomic DNA, extracted after EZL1 or control silencing at a time IES excision is normally finished. In control RNAi experiments, as expected, the 10 IESs analyzed were completely excised from the new developing MACs, and only the rearranged forms, originating from both the old and new MACs, could be detected (Fig. S9). In contrast, IES-retaining forms accumulated in the new MACs of PGM or EZL1 KD

cells, relative to the rearranged sequences present in the fragments of the parental MAC. We observed that EZL1 KD impaired the excision of 7 out of 10 tested IESs, whereas PGM KD impaired the excision of all IESs as expected (Fig. S9).

IESs of sufficient length are circularized after excision (Betermier et al., 2000). We used divergent primers internal to a long IES to monitor the formation of covalently closed excised circles by PCR. No junction products could be detected in Ezl1p-depleted cells, indicating that EZL1 is required for IES circularization (Fig. S6). In addition, *de novo* excision junctions could not be detected in EZL1 KD (not shown). Altogether these data indicate that the EZL1 gene is required for IES excision and, most likely, EZL1p acts upstream of the introduction of DNA double-strand breaks.

Based on our PCR analyses, not all IESs are affected following EZL1 KD. To observe the effects of Ezl1p depletion genome-wide, we used the EZL1 dataset. Using a previously described IES identification pipeline (Arnaiz et al., 2012), 30,325 IESs were found in the EZL1 sample. Of these, 96,6% (29,316) are identical to the genome-wide set of IESs previously identified (Arnaiz et al., 2012) and 99% are found at the same position. A method was developed to measure the retention of each IES (see Materials and Methods). As expected, the retention score of IESs in the control dataset is close to 0 (mean 0.008), whereas a Gaussian distribution was observed in the PGM dataset (Arnaiz et al., 2012) with a mean at 0.77. This suggests that excision of all IESs is affected in a similar manner following PGM KD. In the EZL1 dataset however, the mean of IES retention score is at 0.32 and the distribution, different from that of PGM, is bimodal with 8,085 IESs that have a retention score close to 0 and the rest of the IESs displaying a wide distribution of retention scores (mean 0.32) with a mode of 0.5 (Fig. 5A). Finally, we used stringent statistical tests (see Materials and Methods) to compare the retention scores in the EZL1 or PGM datasets to the retention scores in the control dataset to determine the significantly retained IESs. In the PGM dataset, 44,028 IESs (97.9%) are significantly retained and in the EZL1 dataset, 31,481 IESs (70.1%) are significantly retained. We sequenced a biological replicate obtained following an independent EZL1 silencing experiment and found a good correlation of retention scores (Spearman correlation coefficient: 0.887, $p < 2.2 \times 10^{-16}$) for the two biological replicates. Thus 29,218 IESs were significantly retained in the replicate, while 28,812 IESs were significantly retained in both replicates. Based on these data, we define two classes of IESs: those that are significantly retained after EZL1 KD and those that are not. It is striking that our genome-wide analysis is completely coherent with our PCR analyses (Table S5). In

particular, the 3 IESs that were correctly excised following EZL1 KD in our PCR assay did not show significant retention score in the sequencing data from either biological replicate.

We then wondered what distinguishes the two classes of IESs. Our PCR analyses indicated that long IESs were retained following EZL1 KD (Fig. S9 and Table S5). To confirm this observation genome-wide, IESs were grouped according to their size (Arnaiz et al., 2012) and the distribution of retention scores for each group was calculated and represented in a box plot (Fig. 5B). As expected, the same periodic distribution with peaks every ten base pairs was observed for the two classes of IESs (not shown). The first 5 groups (26-72 pb) have retention score distributions that are significantly different from each other: the larger the IES size, the higher the retention score. Starting with the 5th group (>72 bp), the median does not change significantly, which indicates that IESs of these sizes are similarly affected by Ezl1p depletion. For the largest IESs (>1kb), the retention score distribution is significantly shifted to higher values. Among those, there is one group composed of 28 IESs, which have been shown to derive from Tc1/mariner transposons named Anchois (Arnaiz et al., 2012). All of them are retained after EZL1 inactivation (Table S6). Roughly 50% of IESs are over 52 bp in length and among them, 89.9% are significantly retained after EZL1 KD, while only 40% of the IESs shorter than 52 bp in length are significantly retained. This striking link between IES size and retention score in the EZL1 sample is not observed for the PGM dataset (Fig. S10), indicating that it is thus a property specific to Ezl1p depletion.

We then searched for features other than size that could be associated with EZL1-retained IESs (Fig. S11). We compared IESs of the same size range (26-32 bp) and did not find any meaningful correlation for a large number of criteria, including the consensus present at IES ends and the scnRNA density on IESs. Yet several features were found to be statistically significant in EZL1-retained IESs: (i) a slightly higher GC content, (ii) a more frequent location within gene coding sequences, although the significance of these observations is not clear.

The Dcl2 and Dcl3 proteins are required for the excision of subset of IESs that are highly sensitive to Ezl1p depletion

Since DCL2 and DCL3 genes, like EZL1, are required for establishing H3 K9me3 and K27me3 in the developing MAC (Fig. 2 and 3), we hypothesized that depletion of Dcl2 and Dcl3 proteins would impair DNA elimination in a similar manner to that of Ezl1p

depletion. To address this question, we sequenced DNA isolated from newly developed MACs following DCL2/3 co-silencing. When compared to the PGM and EZL1 datasets, the total complexity of assembled contigs was similar in the DCL2/3 dataset (Table S4) and analysis of Sardine retention showed that all characterized Sardine copies are significantly retained following DCL2/3 co-silencing (Fig. S8). This global analysis confirmed that the Dcl2 and Dcl3 proteins are required for the imprecise elimination of germline-limited sequences (Lepere et al., 2009).

We then analyzed the effects of DCL2/3 KD on IES excision. Compared to PGM or EZL1 silencing, most IESs are weakly or not at all retained after DCL2/3 co-silencing (Fig. 5A). 3,272 IESs (7.3%) are significantly retained in the DCL2/3 dataset with a mean retention score of 0.04. We found 3,160 IESs significantly retained with a mean at 0.05 for a biological replicate (Sandoval et al., 2014). A rather good correlation of IES retention scores was observed for the two biological replicates (Spearman correlation coefficient 0.616, $p < 2.2 \times 10^{-16}$) that were obtained using different silencing constructs. Furthermore, the retention scores measured for the DCL2/3 dataset are in agreement with our PCR analyses (Fig. S9 and Table S5). Based on our PCR analyses, we noticed that all mclESs are significantly retained in the DCL2/3 dataset and that all DCL2/3-retained IESs are retained in the EZL1 dataset. The latter was confirmed genome-wide: all significantly retained IESs (3,269/3,272) in the DCL2/3 dataset are significantly retained in the EZL1 dataset. Furthermore, DCL2/3 significantly retained IESs are among the most retained IESs in the EZL1 sample (Fig. S12). Only the largest IESs are retained in the DCL2/3 dataset. 50% of the IESs larger than 1kb are significantly retained (Fig. S10). Among those, 19/28 Anchois IESs are significantly retained after DCL2/3 inactivation (Table S6). Altogether our data indicate that DCL2/3-retained IESs correspond to a small subset of EZL1-retained IESs (Fig. 5C).

DISCUSSION

Both H3 K9 and K27 tri-methylation appear to participate in the regulatory pathway that controls developmental genome rearrangements in *Paramecium*. Indeed, we observed that: i) K9me3 and K27me3 appear transiently in the developing somatic macronucleus, ii) depletion of the histone methyltransferase Ezl1 abolishes H3 K9me3 and K27me3 and leads to severe defects in genome rearrangements, iii) K9me3, K27me3 and the fusion protein GFP-Ezl1 display a very similar dynamic pattern of localization and both histone

marks colocalize with GFP-Ezl1, iv) both histone marks are greatly reduced in the developing somatic macronucleus after depletion of the Dicer-like 2 and 3 proteins, which are required for scnRNA biogenesis and genome rearrangements (Lepere et al., 2009) and v) depletion of the endonuclease Pgm leads to loss of H3 K9me3 and K27me3 foci in the developing somatic macronucleus.

Alignment of SET domains from various E(z) proteins revealed conservation of key residues implicated in binding the methyl donor, targeting lysine and catalysis, suggesting that the Ezl1 protein is a *bona fide* K27 HMT. The Ezl1 protein might therefore catalyze K27 trimethylation, which in turn would allow K9 trimethylation, as proposed in *Tetrahymena* (Liu et al., 2007). However, it is possible that H3K9 as well as H3K27 methylation is catalyzed by the protein Ezl1p, as no H3K9-specific KMTs homologs could be identified in the somatic genome of ciliates (Fig. S4).

Double substrate specificity has been shown *in vitro* for *Drosophila* E(z) (Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002). Furthermore, several studies support the notion that overlap between the functions of these two modifications may be widespread (Bilodeau et al., 2009; Mozzetta et al., 2014). In ciliates, a single K9me3 and K27me3 pathway appears to contribute to the establishment of repressive chromatin in the developing somatic macronucleus.

The histone methyltransferase EZL1 acts locally and displays a DNA length bias

The work presented here demonstrates that the histone methyltransferase Ezl1 is required for the elimination of transposable elements, of long IESs and of germline DNA regions that encompass several kb in length, which might altogether represent up to one third (25 Mb) of the germline genome. We have shown that H3K9me3 and H3K27me3 signals are abolished after Ezl1p depletion and that scnRNAs are necessary for the deposition of both histone marks in the developing somatic macronucleus. Therefore, our results support the idea that scnRNAs guide the histone methyltransferase Ezl1 to specific germline sequences in the developing somatic macronucleus. Deposition of H3 K27me3 and K9me3 would allow the recruitment, or activity, of the endonuclease Pgm and excision of marked DNA segments (Fig. 6), consistent with our observations that Ezl1 acts upstream of Pgm. Our study provides evidence that RNAi-mediated heterochromatin formation is necessary for elimination of germline DNA in *Paramecium*, as is the case in *Tetrahymena*. In a similar manner to the sRNA mediated heterochromatin formation in *S. pombe* and *Drosophila* (Castel and Martienssen, 2013;

Ge and Zamore, 2013), scnRNAs would guide the heterochromatinization of DNA segments, which would be simply removed by the process of DNA elimination.

One unexpected finding, however, is that the histone methyltransferase Ezh1 is also required for the excision of very short DNA segments, as Ezh1 depletion leads to retention of 31,481 IESs (70,1%). Because the vast majority of IESs are shorter than 150 bp in length, Ezh1 might not necessarily trigger the formation of heterochromatin on eliminated sequences. Instead, we imagine that Ezh1 acts locally and is responsible for trimethylation on lysine 9 and lysine 27 on one or a few nucleosomes that overlap with the IES. The quantitative genome-wide analysis of IES retention after Ezh1 depletion showed that all IESs are not equal: IESs are retained to a different extent after Ezh1 depletion and we could not identify any features in the IES sequences that distinguish IESs that are significantly retained from those that are not. However, our analysis revealed a striking link between IES size and retention score after Ezh1 depletion, as 90% of IESs longer than 52 bp are retained after Ezh1 depletion. We propose that the excision process is regulated by the presence of methylated nucleosomes and depends on the relative positions of IES ends and the methylated nucleosomes. As illustrated on Figure 6, the positioning of nucleosomes might play a major role in IES excision. We reasoned that longer sequences have a higher probability to be associated with modified nucleosomes and would thus be more sensitive to Ezh1 depletion and loss of methylated H3. Any IESs over 52 bp in length would be either entirely or partially covered by one nucleosome. This might reflect the length of the linker DNA in between nucleosomes in the developing somatic macronucleus, which is not known, but would be consistent with linker ranging from 20 to 90 bp in general. Strikingly, small IESs between 26 and 52 bp in length have variable retention scores. 34 % of the smallest IESs (26-32 bp) required Ezh1 to be excised and we imagine those small IESs are either within, or partially covered by, one modified nucleosome. IESs that are not retained after Ezh1 depletion might be located in the linker region in between nucleosomes and alternative mechanisms would ensure their correct and precise excision. In *S. cerevisiae*, the chromatin remodeler SWR1 binds *in vitro* long nucleosome-free DNA and the adjoining nucleosome core particle, allowing discrimination of gene promoters over gene bodies. SWR1 binding is enhanced on acetylated nucleosomes, but recognition of nucleosome-free and nucleosomal DNA is dominant over interaction with acetylated histones (Ranjan et al., 2013). Such hierarchical cooperation between DNA and posttranslational histone modifications might participate in guiding the excision machinery. Precise mapping of

nucleosomes and of histone marks along the genome will be needed to explore this possibility. An exciting challenge for the future is to understand the mechanisms by which differential markings by histone modifications define positions of the excision machinery and precise DNA cleavage sites.

Different classes of IESs

One other important finding is the existence of different classes of IESs. While all of them are ultimately excised in a Pgm-dependent manner (Arnaiz et al., 2012), IESs appear to differ in their recognition mechanism. Only about a third of IESs (5 out of 13 tested), called mcIESs, are sensitive to the presence of homologous sequences in the maternal MAC, while this does not affect the excision of non-mcIESs (Duharcourt et al., 1998). Interestingly, genome-wide analyses of the effects of depletion of Dicer-like 2 and 3 proteins showed that they are both required for Tc1/mariner TEs and of mcIES, but not of non-mcIESs (Sandoval et al., 2014 and this study). The evidence obtained so far is consistent with the idea that Dcl2-3 retained IESs correspond to mcIESs, but unfortunately, it is not possible to experimentally determine the genome-wide set of mcIESs. Surprisingly however, only a small fraction of IESs (less than 10%) are retained after depletion of the Dicer-like 2 and 3 proteins, indicating that the fraction of mcIESs in the genome might be smaller than initially thought.

More importantly, our data indicate that most IESs are correctly excised in the absence of scnRNAs. IESs, even those that are not under maternal control, normally do produce scnRNAs during MIC meiosis (Lepere et al., 2009; Sandoval et al., 2014; Singh et al., In press) and, when introduced into the maternal MAC, give rise to non-coding transcripts like any other sequence (Lepere et al., 2008), suggesting the genome scanning process should inactivate their scnRNAs. Our conclusion is thus that excision of non-mcIESs simply does not depend on scnRNAs. None of the shortest (~28 bp) IESs tested, which are also the oldest, was found to be a mcIES, raising the possibility that non-mcIES represent the endpoint of IES evolution. In support of this view, our data indicate that recognizable TEs and young (longer) IESs display higher retention score after depletion of Dicer-like 2 and 3 proteins, indicating that they indeed depend on their own scnRNAs for recognition and elimination.

In addition to these two classes of IESs, genome wide analysis of the effects of EZL1 depletion provided evidence for additional classes of IESs, showing differential sensitivity to the different factors studied here. In order to group IESs into functionally similar

classes, we have quantitated the requirement of each of the 45,000 IESs (and of known transposons) for each of the factors analyzed (Figure 5C). Our data showed that the vast majority of IESs are retained after EZL1 depletion, suggesting that both H3K9me3 and H3K27me3 are required for IES excision as discussed above. One surprising finding is that the set of EZL1 retained IESs does not entirely overlap with that of DCL2-3 retained IESs. Even though DCL2-3 retained IESs are all included in EZL1 retained IESs, our results argue that EZL1 is necessary for correct excision of most IESs, without the need of scnRNAs. This observation raises the possibility that H3K9me3 and H3K27me3 methylation may be loaded on IESs independently of scnRNAs. Because the excision of IESs, whether they are maternally controlled or not, EZL1 sensitive or not, is still dependent on the Ptiwi01/09 proteins (Bouhouche et al., 2011), we propose that these proteins may be alternatively loaded with a different type of small RNA not subjected to genome scanning.

With our present work, we now understand that programmed genome rearrangements involve at least three partially overlapping pathways. Whether different protein complexes or the same complex whose individual components in combination with genomic sequence, nucleosome organization and other not yet identified features contribute to these different pathways is not yet known. Future studies combining genetic and biochemical approaches will help reveal the level of complexity that is now beginning to emerge.

MATERIALS AND METHODS

***Paramecium* strains, cultivation and autogamy**

Unless otherwise stated, all experiments were carried out with the entirely homozygous strain 51 of *P. tetraurelia*. Cells were grown in wheat grass powder (WGP) (Pines International) infusion medium bacterized the day before use with *Klebsiella pneumoniae*, unless otherwise stated, and supplemented with 0.8 mg/mL β -sitosterol (Merck). Cultivation and autogamy were carried out at 27°C as described (Beisson et al., 2010a, 2010b).

Gene silencing experiments

Plasmids used for T7Pol-driven dsRNA production in silencing experiments were obtained by cloning PCR products from each gene using plasmid L4440 and *Escherichia coli* strain HT115 DE3, as previously described (Galvani and Sperling, 2002). Sequences used for silencing of EZL2, EZL3a, EZL3b, EZL4 were segments 955-1519; 1402-1980; 1404-1982; and 1398-1976 of GSPATG00032888001; GSPATG00012695001; GSPATG00013305001; PTETG1700020001 (see ParameciumDB at <http://paramecium.cgm.cnrs-gif.fr/>), respectively. For EZL1 silencing, two non-overlapping gene fragments covering positions 991-1500 (EZL1-1) and 332-754 (EZL1-2) of GSPATG00037872001 were used. The fragments used for ND7, ICL7a (Bouhouche et al., 2011), DCL2, DCL3 (Lepere et al., 2009) and PGM-1 (Baudry et al., 2009). Preparation of silencing medium and RNAi during autogamy were performed as described in Baudry et al., 2009. Lethality of post-autogamous cells after double silencing of DCL2 and DCL3 or silencing of PGM was 90-100% (30-60 cells were checked in each silencing experiment).

Injection of GFP fusion transgene

For the construction of in-frame GFP-EZL1 fusion, a GFP-coding fragment adapted to *Paramecium* codon usage (Nowacki et al., 2005) was added by PCR fusion to the 5' end of the EZL1 gene. As a result, the GFP is fused to the N-terminus of EZL1 and the fusion protein is expressed under the control of the EZL1 transcription signals (promoter and 3'UTR). It contains the 830-bp genomic region upstream of the EZL1 open reading frame, the 304-bp genomic region downstream.

Plasmid carrying the GFP-EZL1 fusion transgene was linearized by XmnI and microinjected into the MAC of vegetative 51 cells. No lethality was observed in the post-autogamous progeny of injected cells (data not shown), indicating that the GFP-EZL1 fusion did not interfere with normal progression of autogamy (Fig. 3 and S7).

DNA extraction, Southern blot and PCR

DNA samples were typically extracted from 200-400-ml cultures of exponentially growing cells at <1000 cells/ml or of autogamous cells at 2000-4000 cells/ml as previously described (Beisson et al., 2010b). Small-scale DNA samples were prepared from $\leq 1,000$ cells using the NucleoSpin® Tissue kit (Macherey-Nagel). Electrophoresis and blotting were carried out according to standard procedures.

PCR amplifications were performed in a final volume of 25 μ L, with 10 pmol of each primer, 10 nmol of each dNTP and 1.9 U of Expand Long Template Enzyme mix (Expand Long Template PCR system, Roche). PCR products were analyzed on 0.8% agarose gel (Fig. 4). Oligonucleotides were purchased from Eurofins MWG Operon (see Table S7).

Histone extraction and western blot

Cell pellets were mechanically lysed in three volumes of lysis solution (0.25 M sucrose, 10 mM MgCl₂, 10 mM Tris pH 6.8, 0.2% Nonidet P-40) with a Potter-Elvehjem homogenizer. Following the addition of 2.5 volumes of washing solution (0.25 M sucrose, 10mM MgCl₂, 10 mM Tris pH7.4), the nuclei-containing pellet was collected by centrifugation at 1000 g for 1 min. The histone samples were-suspended in 0.4 N sulphuric acid and incubated on wheel overnight at 4°C. After centrifugation at 16,000 g for 10 min, the acid-extracted histones were precipitated by TCA (20% w/v). After washing once with cold acetone, the histone samples were re-suspended in water prior to quantification with the Qubit® Protein Quantification Assay (Invitrogen, Life Technologies Corporation).

10 μ g of histone extracts were used for western blot. Electrophoresis and blotting were carried out according to standard procedures. The H3K27me3 (1:500; Millipore, 07-449) and H3 (1:10 000; Millipore, 07-690) primary antibodies were used. Secondary horseradish peroxidase-conjugated donkey anti-rabbit IgG antibody (Promega) was used at 1:3000 dilution followed by detection by ECL (SuperSignal West Pico Chemiluminescent Substrate, Thermo Scientific). For normalization, the membranes

probed with H3K27me3 antibody were stripped in mild stripping buffer (glycine 200mM, SDS 0.1%, Tween-20 1%, pH 2.2) and probed again with H3 antibody.

Indirect immunofluorescence and fluorescence quantification.

Paramecium cells were fixed for 30 minutes in solution I (PHEM 1X, paraformaldehyde 1%, Triton X-100 2.5%, Sucrose 4%) and for 10 minutes in solution II (PHEM 1X, paraformaldehyde 6.5%, Triton X-100 1.2%, Sucrose 4%). The primary antibodies used were rabbit polyclonal α -H3K9me3 (07-442, Millipore) and α -H3K27me3 (07-449, Millipore) at 1:500. After overnight incubation with the primary antibodies, cells were washed twice in 1X phosphate-buffered saline (PBS) and incubated with the secondary antibodies (Alexa Fluor 568-conjugated goat anti-rabbit IgG, A-11036, Invitrogen) at 1:500 for 1h. After two washes in 1X PBS, cells were stained with 2 μ g/mL Hoechst, washed once in 1X PBS, centrifuged on microscope slides with the CytoSpin™ 4 Cytocentrifuge (Thermo Scientific) and finally mounted in Citifluor AF2 glycerol solution (Citifluor Ltd, London). Images were acquired using a Zeiss LSM 710 laser-scanning confocal microscope and a Plan-Apochromat 63x/1.40 oil DIC M27 objective. Z-series were performed with Z-steps of 0.5 μ m.

To quantify the H3K9me3 and H3K27me3 signals, the Imaris 3D visualization software (Bitplane) was used. For each time point, the fluorescence intensities of H3K9me3 (or H3K27me3) in the developing MACs (signal) and in the corresponding volume of the cytoplasm (noise) were measured. The mean value and standard deviation of the signal to noise ratios were calculated using ten individual cells at each time point.

DNA sequencing

DNA for deep-sequencing was isolated from post-autogamous cells as previously described (Arnaiz et al., 2012) and sequenced by a paired-end strategy using Illumina GA-IIx and Hi-Seq next-generation sequencers (Table S3).

Reference Genomes

The following reference genomes (Arnaiz et al., 2012) were used in the IES analyses and for read mapping.

MAC reference (strain 51):

http://paramecium.cgm.cnrs-gif.fr/download/fasta/ptetraurelia_mac_51.fa

MAC+IES reference (strain 51):

http://paramecium.cgm.cnrs-gif.fr/download/fasta/ptetraurelia_mac_51_with_ies.fa

PGM contigs:

http://paramecium.cgm.cnrs-gif.fr/download/fasta/assemblies/ptetraurelia_PGM_k51_ctg.fa

Macronuclear DNA reads for PiggyMac depleted cells and for a biological replicate of the DCL2/3 co-silencing experiment (Sandoval et al., 2014) were obtained from the European Nucleotide Archive (Accession number ERA137420) and the GenBank Sequence Read Archive (Accession number SRX387766), respectively.

Genome-wide analysis of IES retention

After quality filtering and removal of adapters, Illumina reads were aligned to the reference genomes (*P. tetraurelia* MAC reference genome and MAC+IES reference genome) using BWA (Li and Durbin, 2009) with default parameters. Alignments were indexed with Samtools (Li et al., 2009).

For each sample, IES retention scores were determined as follows, for each IES in the genome previously identified in Arnaiz et al., 2012. The number of reads that contain the IES sequence (symbolised IES⁺) and the number of reads that contain only the macronuclear IES junction consisting of a TA dinucleotide (IES⁻) were determined. Only reads with unambiguous alignments were counted. Each read was counted only once to avoid over-counting owing to paralogous matches. Reads were only counted at IES ends, to avoid length biases resulting from IES length variation. The retention score (RS) of an IES is then given by the following equation:

$$RS = (IES^+) / (IES^+ + IES^-)$$

To ensure that the observed retention of an IES in an experiment is owing to the gene silencing, we compared its RS to the one observed in the control. First we calculated the confidence interval (alpha = 0.95) of the control retention score value, using the Pearson-Klopper exact method as implemented by the R binom package version 1.0-5 (Dorai-Raj, 2009). Then we tested for higher retention in the experiment, thanks to a frequency comparison test (based on a binomial law of probability) between the experimental retention score and the upper bound of the confidence interval in the control. Resulting p-values were adjusted for multiple testing using the Benjamini & Hochberg method (Benjamini and Hochberg, 1995). IESs with adjusted p-value < 0.05 are considered significantly retained in the sample.

ACKNOWLEDGMENTS

We thank all lab members for continuous support and discussions. We thank Olivier Arnaiz, Laurent Duret and Franck Picard for advice on data analysis, and Sophie Malinsky for provision of the PiggyMac silencing construct. This work was supported by the ANR-2010-BLAN-1603 “GENOMAC”, the ANR-12-BSV6-0017 “INFERNO”, a CNRS ATIP-Plus grant to SD. MLA, AF, CDW were supported by Ph.D. fellowships from the Ministère de l’Enseignement Supérieur et de la Recherche, and MLA was recipient of a fellowship from the Fondation de la Recherche Médicale. We acknowledge the ImagoSeine facility, member of the France BioImaging infrastructure supported by the ANR-10-INSB-04. The sequencing benefited from the facilities and expertise of the high-throughput sequencing platform of IMAGIF (Centre de Recherche de Gif - www.imagif.cnrs.fr). This study was carried out in the context of the CNRS-supported European Research Group “Paramecium Genome Dynamics and Evolution”, the European COST Action BM1102 and the LABEX Who am I? supported by the ANR-11-LABX-0071_WHOAMI and the ANR-11-IDEX-0005-02.

REFERENCES

- Arnaiz, O., Gout, J.F., Betermier, M., Bouhouche, K., Cohen, J., Duret, L., Kapusta, A., Meyer, E., and Sperling, L. (2010). Gene expression in a paleopolyploid: a transcriptome resource for the ciliate *Paramecium tetraurelia*. *BMC Genomics* **11**, 547.
- Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.-M., Wilkes, C.D., Garnier, O., Labadie, K., Lauderdale, B.E., Le Mouël, A., et al. (2012). The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet.* e1002984.
- Aury, J.M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N., et al. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**, 171–178.
- Baudry, C., Malinsky, S., Restituto, M., Kapusta, A., Rosa, S., Meyer, E., and Betermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev* **23**, 2478–2483.
- Beisson, J., Betermier, M., Bre, M.H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., et al. (2010a). Maintaining clonal *Paramecium*

tetraurelia cell lines of controlled age through daily reisolation. *Cold Spring Harb Protoc* 2010, pdb prot5361.

Beisson, J., Betermier, M., Bre, M.H., Cohen, J., Duhaucourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., et al. (2010b). Mass culture of *Paramecium tetraurelia*. *Cold Spring Harb Protoc* 2010, pdb prot5362.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing; *J. R. Stat. Soc. Ser. B* 57, 289–300.

Berger, J.D. (1986). Autogamy in *Paramecium*. Cell cycle stage-specific commitment to meiosis. *Exp Cell Res* 166, 475–485.

Betermier, M., Duhaucourt, S., Seitz, H., and Meyer, E. (2000). Timing of developmentally programmed excision and circularization of *Paramecium* internal eliminated sequences. *Mol Cell Biol* 20, 1553–1561.

Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., and Young, R.A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* 23, 2484–2489.

Bouhouche, K., Gout, J.F., Kapusta, A., Betermier, M., and Meyer, E. (2011). Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res* 39, 4249–4264.

Castel, S.E., and Martienssen, R.A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.* 14, 100–112.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* 111, 185–196.

Dorai-Raj, S. (2009). binom: Binomial Confidence Intervals For Several Parameterizations. R package version 1.0-5. <http://CRAN.R-project.org/package=binom>.

Duhaucourt, S., Keller, A.M., and Meyer, E. (1998). Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. *Mol Cell Biol* 18, 7075–7085.

Galvani, A., and Sperling, L. (2002). RNA interference by feeding in *Paramecium*. *Trends Genet* 18, 11–12.

Garnier, O., Serrano, V., Duhaucourt, S., and Meyer, E. (2004). RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*.

Mol Cell Biol 24, 7370–7379.

Ge, D.T., and Zamore, P.D. (2013). Small RNA-Directed Silencing: The Fly Finds Its Inner Fission Yeast? *Curr. Biol.* 23, R318–R320.

Gogendeau, D., Klotz, C., Arnaiz, O., Malinowska, A., Dadlez, M., de Loubresse, N.G., Ruiz, F., Koll, F., and Beisson, J. (2008). Functional diversification of centrins and cell morphological complexity. *J. Cell Sci.* 121, 65–74.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* 16, 2893–2905.

Lepere, G., Betermier, M., Meyer, E., and Duhaucourt, S. (2008). Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev* 22, 1501–1512.

Lepere, G., Nowacki, M., Serrano, V., Gout, J.F., Guglielmi, G., Duhaucourt, S., and Meyer, E. (2009). Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res* 37, 903–915.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* 25, 2078–2079.

Liu, Y., Mochizuki, K., and Gorovsky, M.A. (2004). Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in *Tetrahymena*. *Proc Natl Acad Sci U A* 101, 1679–1684.

Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in *Tetrahymena*. *Genes Dev* 21, 1530–1545.

Madireddi, M.T., Coyne, R.S., Smothers, J.F., Mickey, K.M., Yao, M.C., and Allis, C.D. (1996). Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in *Tetrahymena*. *Cell* 87, 75–84.

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., and Ait-Si-Ali, S. (2014). The Histone H3 Lysine 9 Methyltransferases G9a and GLP Regulate Polycomb Repressive Complex 2-Mediated Gene Silencing. *Mol. Cell.*

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L.,

O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* 111, 197–208.

Nowacki, M., Zagorski-Ostojka, W., and Meyer, E. (2005). Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in *Paramecium tetraurelia*. *Curr Biol* 15, 1616–1628.

Nowak, J.K., Gromadka, R., Juszczuk, M., Jerka-Dziadosz, M., Maliszewska, K., Mucchielli, M.H., Gout, J.F., Arnaiz, O., Agier, N., Tang, T., et al. (2011). A functional study of genes essential for autogamy and nuclear reorganization in *Paramecium*. *Eukaryot Cell*.

Ranjan, A., Mizuguchi, G., FitzGerald, P.C., Wei, D., Wang, F., Huang, Y., Luk, E., Woodcock, C.L., and Wu, C. (2013). Nucleosome-free region dominates histone acetylation in targeting SWR1 to promoters for H2A.Z replacement. *Cell* 154, 1232–1245.

Sandoval, P.Y., Swart, E.C., Arambasic, M., and Nowacki, M. (2014). Functional Diversification of Dicer-like Proteins and Small RNAs Required for Genome Sculpting. *Dev. Cell* 28, 174–188.

Singh, D.P., Saudemont, B., Guglielmi, G., Arnaiz, O., Goût, J.-F., Prajer, M., Potekhin, A., Przybòs, E., Aubusson, A., Aubusson-Fleury, A., et al. (In press). Genome-defence small RNAs exapted for epigenetic mating-type inheritance.

FIGURE LEGENDS

Fig. 1. Dynamic localization of H3 K27me3 and K9me3 in the developing somatic macronuclei (MAC).

Immunofluorescence of cells at different stages of nuclear differentiation are shown, aligned with their schematic representations on the left: (a) T10 hrs; (b) T20 hrs; (c) T30 hrs; (d) T40 hrs. T0 hr corresponds to 50% of cells with fragmented old MAC and the time-points refer to hours after T0. Overlay of Z-projections of magnified views of H3K27me3- or H3K9me3- specific antibodies (in green), and Hoechst (in red) on selected stacks are presented. Dashed white circles indicate the two developing macronuclei and the micronuclei when visible. The other Hoechst-stained nuclei are fragments from the old vegetative MAC. The scale bar is 5 μ m. See Figure S2 for the entire images and a description of the staining throughout the life cycle. See Figure S3A for details on the progression of autogamy and quantification of the number of stained cells at each time point.

Fig. 2. Depletion of the Pgm endonuclease and of the Dicer-like 2 and 3 proteins alter H3K9me3 and H3K27me3 localization.

Immunolabeling with H3K27me3- or H3K9me3- antibodies (in green) and staining with Hoechst (in red) **(A)** in ND7 (control) or PGM KD cells at T30 hrs, and **(C)** in ND7 (control) or DCL2/3 KD cells at T6 hrs. See Figure S3 for progression of autogamy. Dashed white circles indicate the two developing macronuclei and the micronuclei when visible. Scale bar is 5 μ m. Acid extracts from nuclei isolated at T30 hrs from control or PiggyMac-depleted **(B)** and control or DCL2/3- depleted cells **(D)** were resolved on 15% SDS-PAGE, blotted, and probed with the indicated antibodies. Quantification of H3K27me3 signal was normalized with H3 (K27me3/H3) and shown on the graph below.

Fig. 3. EZL1 is required for H3K9me3 and H3K27me3 in the developing somatic MAC.

A. GFP-EZL1 localization in the developing new somatic MAC. Overlay of Z-projections of magnified views of GFP-EZL1 (in green) and Hoechst (in red) on selected stacks at

different stages of development of the somatic MAC (a-e) are presented. See Figure S5 for the entire images and a description of the staining throughout the life cycle. Dashed white circles indicate the two developing macronuclei and the micronuclei when visible. Scale bar is 5 μ m.

B. Immunolabeling with H3K27me3- or H3K9me3- antibodies (in green) and staining with Hoechst (in red) in ND7 (control) or EZL1 knockdown cells at T30 hrs. See Figure S3 for progression of autogamy and quantification of the number of stained cells at each time point. Dashed white circles indicate the two developing macronuclei. Scale bar is 5 μ m.

Quantification of H3K9me3 (**C**) or H3K27me3 fluorescence intensities (**D**) in the developing MAC. At each time point, the mean of fluorescence intensities was calculated for ten individual cells after control or EZL1 silencing (see Materials and Methods). Errors bars represent the standard deviation.

E. Acid extracts from purified nuclei isolated at T30 hrs were resolved on 15% SDS-PAGE, blotted, and probed with the indicated antibodies.

Fig. 4. EZL1 is required for imprecise DNA elimination.

A. Fragmentation of germline DNA downstream of the G51 gene is analyzed by Southern blot hybridization of PstI-digested total genomic DNA run on 0.8% agarose gel after ICL7 (control) or EZL1 KD. ICL7 is a non essential gene that encodes an infraciliary lattice centrin (Gogendeau et al., 2008). The subtelomeric tel51G probe (Baudry, 2009) is shown as a grey rectangle above the line. The black square represents telomeric repeats of the MAC chromosome.

B. Maternal inheritance of macronuclear deletion is analyzed after ICL7 (control) or EZL1 RNAi, using a cell line reproducibly deleting the ND7 gene from the MAC genome at each sexual generation. PCR analysis was performed on the same DNA samples than in (A) with primers (black arrows) located upstream and downstream the ND7 open reading frame.

Fig. 5. Genome-wide effects of EZL1 and DCL2/3 KD on IES excision.

A. IES retention score after PGM, EZL1 or DCL2/3 KD. Histograms of IES retention scores, as defined in Materials and Methods, for control (grey), PGM (red) and EZL1 (green) silencing and DCL2/3 co-silencing (yellow) datasets.

B. EZL1 retention score displays a strong bias for IES size. IESs were grouped by size, each group corresponding to a peak in the genome-wide IES size distribution (Arnaiz et al., 2012). Each box summarizes the EZL1 retention score distribution for the group. The median EZL1 retention score (horizontal line inside the box) and the first (top of box) and third (bottom of box) quantiles are shown. Stars beneath the median value of a group indicate that the retention score distribution of that group is significantly different from the retention score distribution of the previous group. One star, $p < 0.05$, two stars, $p < 1e-10$, three stars, $p < 2.2e-16$. The groups contain: 15857; 485; 6354; 3514; 3108; 3642; 2459; 1934; 1532; 1111; 795; 594; 410; 1176; 780; 398; 198; 89; 64; 47; 183; 63; 135 IESs, respectively.

C. Venn diagram of significantly retained IESs after PGM, EZL1 or DCL2/3 silencing.

Fig. 6. Model for the action of the histone methyltransferase EZL1 in programmed genome rearrangements.

To take into account the length bias of DNA sequences retained after Ezl1 depletion, we propose that efficient excision of germline-limited DNA segments (orange) is regulated by the position of methylated nucleosomes. For long DNA segments (>150 bp in length), that are covered by at least one nucleosome, the histone methyltransferase EZL1 would be targeted to the eliminated sequences and would catalyze the trimethylation of H3K27 and, directly or not, the trimethylation of H3K9. These histone marks would attract the excision machinery and allow the excision of marked DNA segments. DNA segments whose size is comprised between 52 bp and 150 bp would be partly or entirely included into one nucleosome and histone H3 methylation would be essential for their recognition and excision. The smallest DNA segments (26-52 bp in length), however, would be either wrapped around a nucleosome and histone methylation might be needed for efficient excision, or located within the linker DNA in between nucleosomes and histone methylation might be dispensable for their excision.

Figure 1

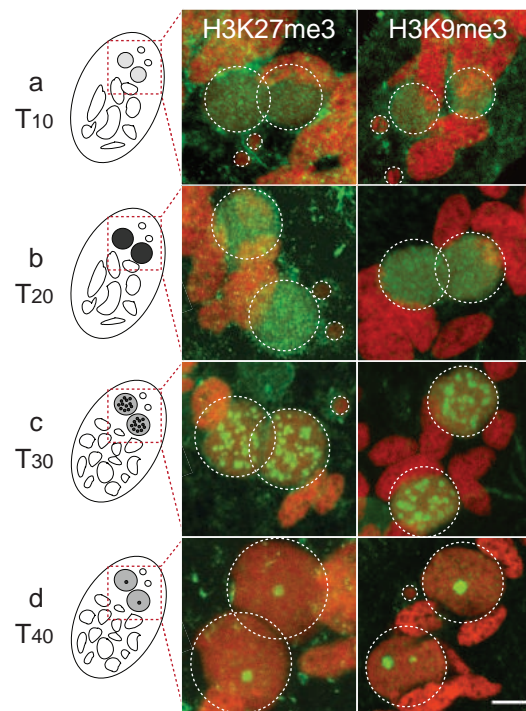


Figure 2

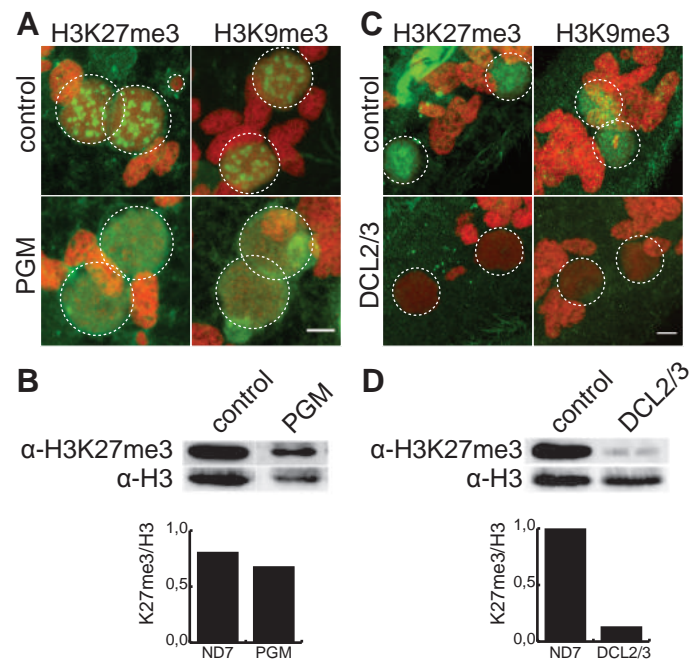


Figure 3

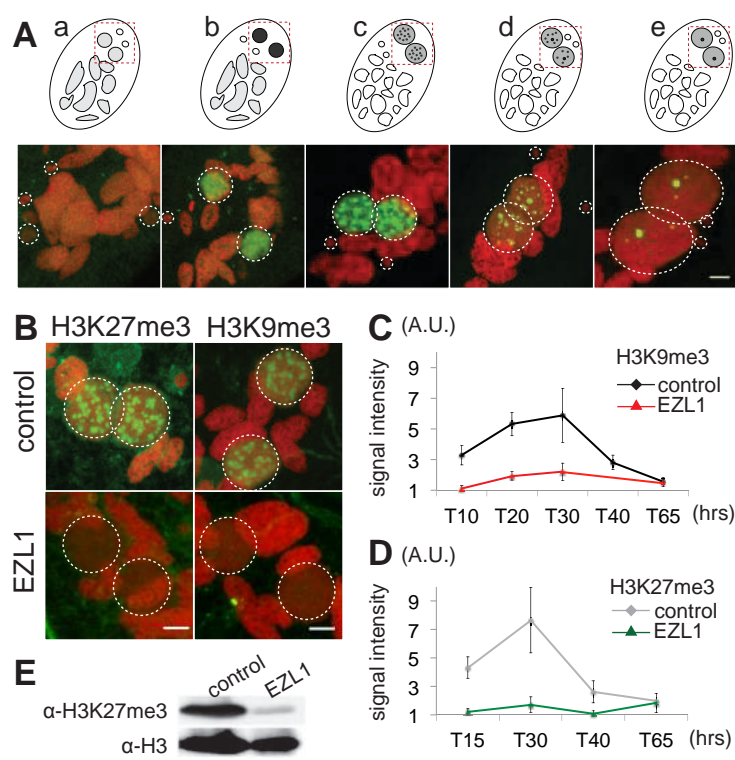


Figure 4

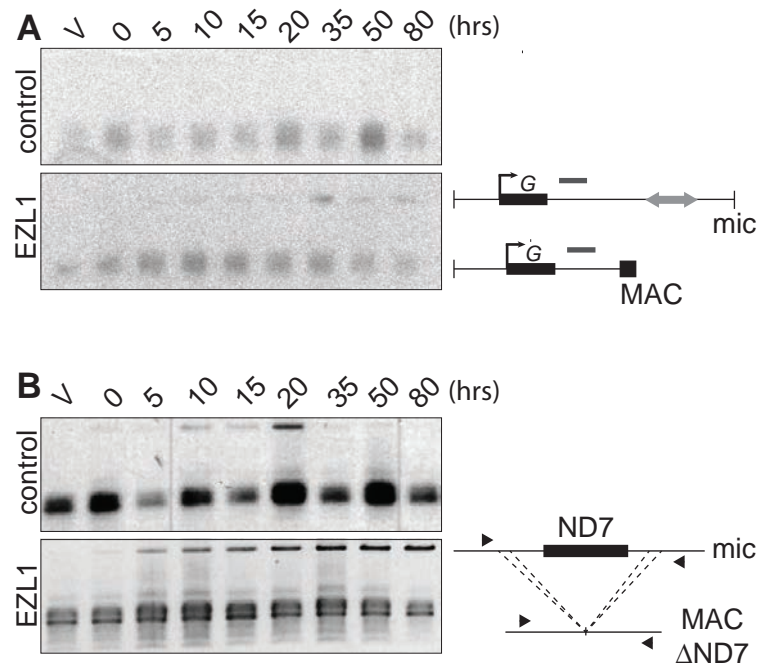


Figure 5

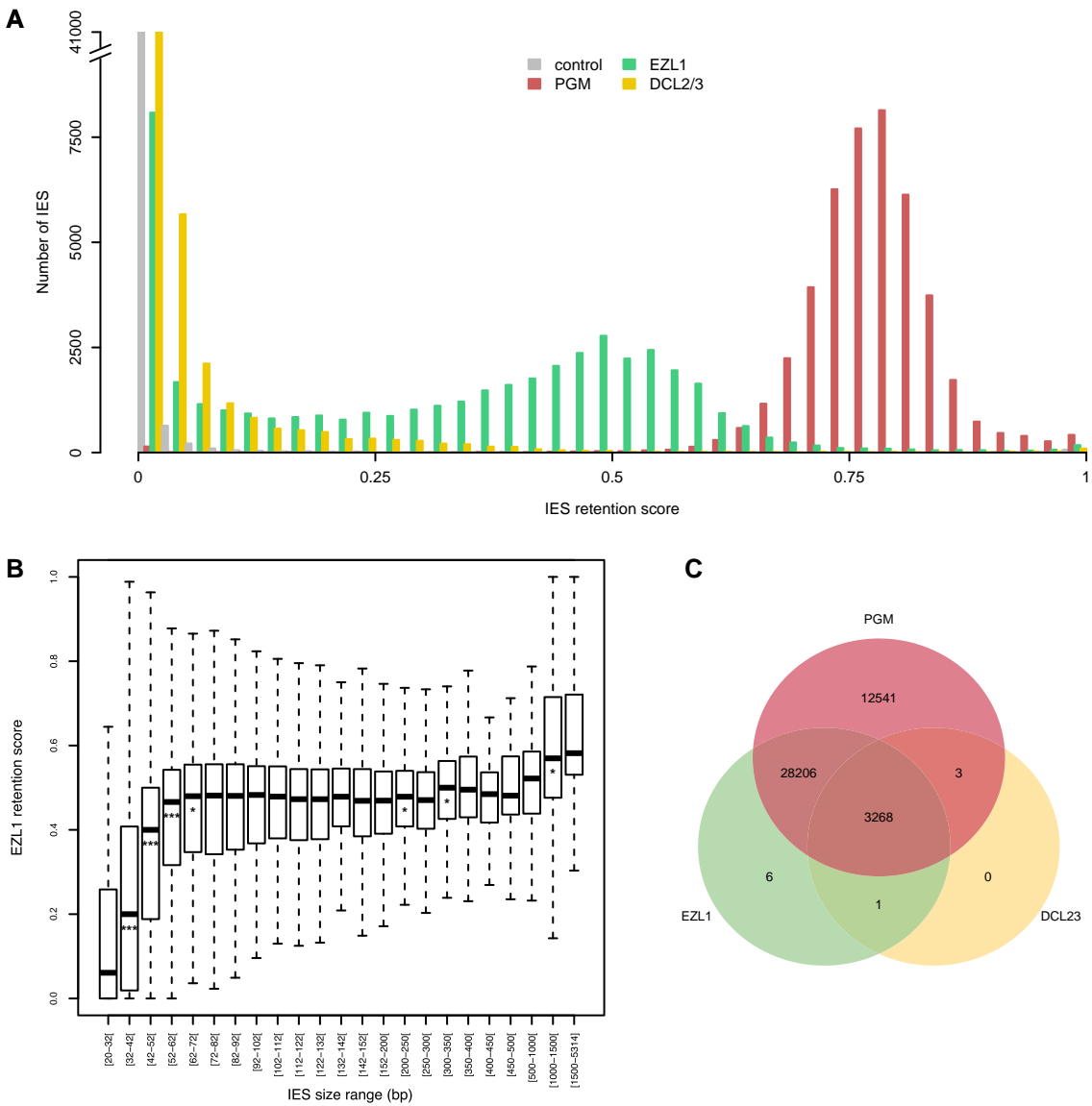
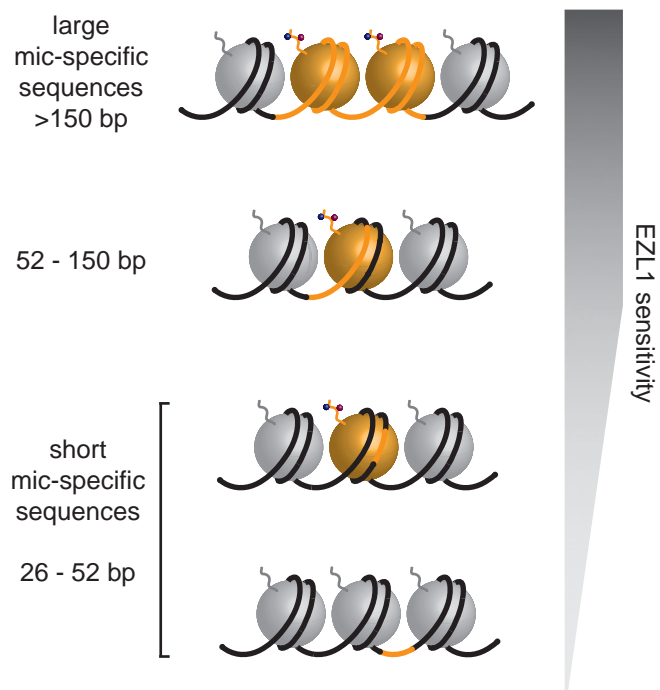


Figure 6



SUPPLEMENTARY MATERIAL

SUPPLEMENTARY MATERIALS AND METHODS

Phylogenetic analyses

SET domain proteins were retrieved using Pfam (Punta et al., 2012) and BLAST (Altschul et al., 1990). Multiple alignments were performed with MUSCLE 3.8 (Edgar, 2004) and were subsequently manually improved. Maximum likelihood (ML) analyses were performed with PHYML (Guindon and Gascuel, 2003) using the PHYML web server (Guindon et al., 2005) hosted at the Montpellier bioinformatics platform (<http://www.atgc-montpellier.fr/phyml/>). PHYML analyses were performed using the Le and Gascuel (LG) amino-acid substitution model (Le and Gascuel, 2008), using two rate categories (one constant and four γ rates). Statistical support for the different internal branches was assessed by approximate Likelihood-ratio test (aLRT; Anisimova and Gascuel, 2006).

RNA extraction, RT-PCR and PCR

RNA samples were typically extracted from 200-400-ml cultures of exponentially growing cells at <1000 cells/ml or of autogamous cells at 2,000-4,000 cells/mL as previously described (Beisson et al., 2010). RNA samples were reverse-transcribed with RevertAid H Minus Reverse Transcriptase (Thermo Scientific) using polydT primers (Thermo Scientific) according to the manufacturer's instructions. It was then followed by PCR amplifications in a final volume of 25 μ L, with 10 pmol of each primer, 10 nmol of each dNTP and 2 U of DyNAzyme II DNA polymerase (Thermo Scientific). For PCR analysis of IES excision (**Figures S4 and S8**), PCR amplifications were performed with 1.9 U of Expand Long Template Enzyme mix (Expand Long Template PCR system, Roche). Oligonucleotides were purchased from Eurofins MWG Operon (see **Table S5**).

sRNA-seq Samples

sRNA-seq samples (Sandoval et al. 2014) were recovered from the GenBank Short Read Archive (Accession No. SRR907882- SRR907875), representing sRNA-Seq data for early and late time points (cf. Sandoval et al. 2014) from DCL5 silenced cells and early and late time points from control cells. The DCL5 silenced samples were used for analysis of 25 nt scnRNAs since in the absence of Dcl5p, biogenesis of iesRNAs, some of which are 25 nt in size, is completely abolished. The control sample (late time point)

was used for analysis of 26-29 nt iesRNAs.

sRNA-seq Mapping

After removal of Illumina sequencing adaptors, reads were discarded if they matched known contaminants (bacterial DNA, rDNA, mitochondrial DNA, plasmid DNA). Reads were separated into different size classes (15 – 35 nt) and mapped with BWA (version 0.6.2-r126) (Li and Durbin, 2009) using default parameters except that 0 mismatches were tolerated.

sRNA Normalization

To allow comparison of the different sRNA populations, the read counts were normalised with respect to the total number of reads mapped on the MAC and the MAC+IES reference genomes between 15 and 35 nt in each sample.

SUPPLEMENTARY REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552.

Arnaiz, O., Gout, J.F., Betermier, M., Bouhouche, K., Cohen, J., Duret, L., Kapusta, A., Meyer, E., and Sperling, L. (2010). Gene expression in a paleopolyploid: a transcriptome resource for the ciliate *Paramecium tetraurelia*. *BMC Genomics* 11, 547.

Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.-M., Wilkes, C.D., Garnier, O., Labadie, K., Lauderdale, B.E., Le Mouël, A., et al. (2012). The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet.* 8, e1002984.

Aury, J.M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N., et al. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444, 171–178.

Beisson, J., Betermier, M., Bre, M.H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., et al. (2010). Mass culture of *Paramecium tetraurelia*. *Cold Spring Harb Protoc* 2010, pdb prot5362.

Bouhouche, K., Gout, J.F., Kapusta, A., Betermier, M., and Meyer, E. (2011). Functional

specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res* 39, 4249–4264.

Duharcourt, S., Keller, A.M., and Meyer, E. (1998). Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. *Mol Cell Biol* 18, 7075–7085.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005). PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 33, W557–559.

Le, S.Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–1320.

Punta, M., Coghill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40, D290–301.

Sandoval, P.Y., Swart, E.C., Arambasic, M., and Nowacki, M. (2014). Functional Diversification of Dicer-like Proteins and Small RNAs Required for Genome Sculpting. *Dev. Cell* 28, 174–188.

Sun, X.-J., Xu, P.-F., Zhou, T., Hu, M., Fu, C.-T., Zhang, Y., Jin, Y., Chen, Y., Chen, S.-J., Huang, Q.-H., et al. (2008). Genome-wide survey and developmental expression mapping of zebrafish SET domain-containing genes. *PloS One* 3, e1499.

Zhang, L., and Ma, H. (2012). Complex evolutionary history and diverse domain organization of SET proteins suggest divergent regulatory interactions. *New Phytol.* 195, 248–263.

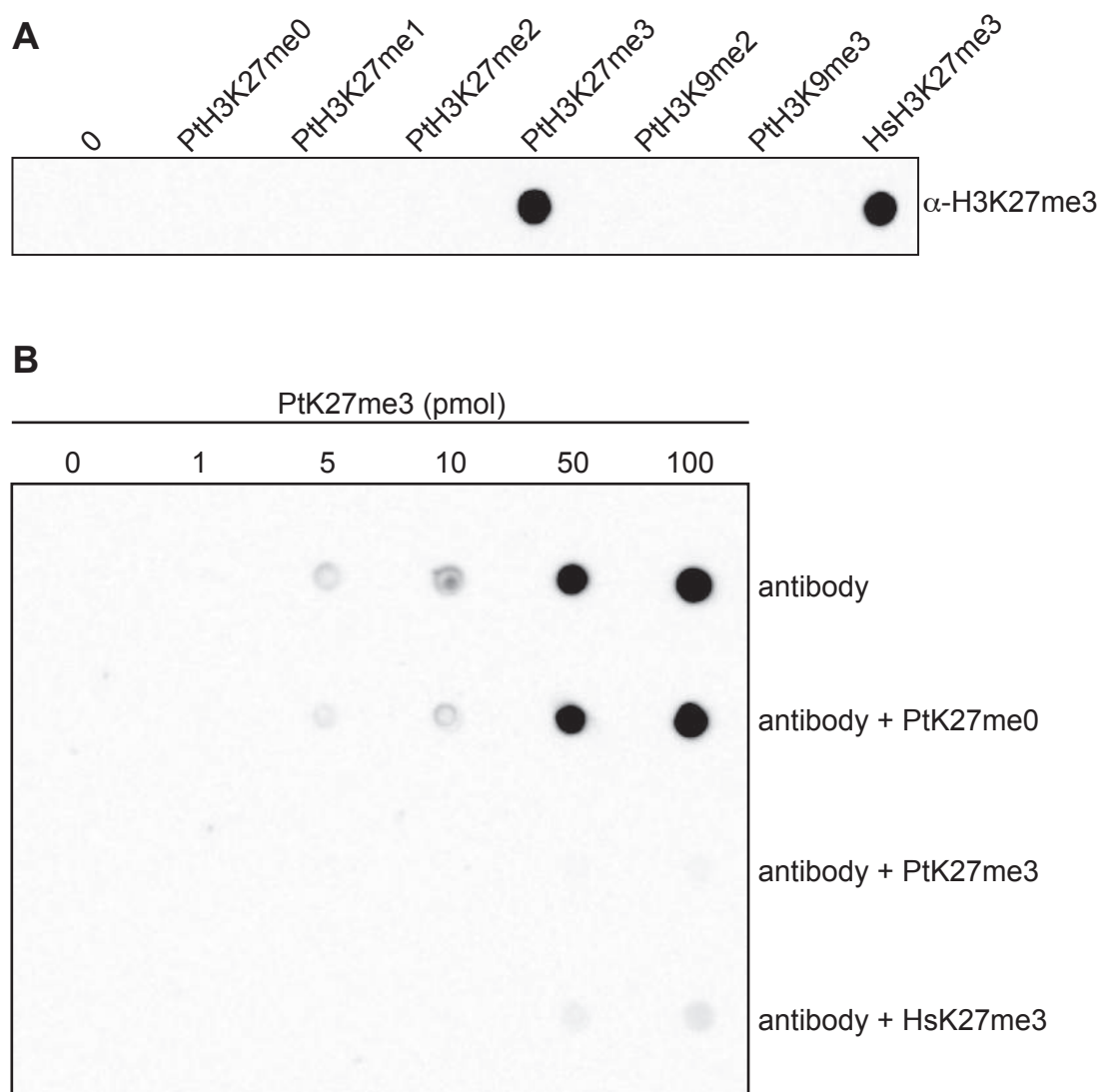


Figure S1. Specificity of commercial H3K27me3 antibodies in *Paramecium*.

A. Dot blot assay using H3 peptides. 100 pmol of each of the indicated peptide was spotted on the membrane, and probed with the H3K27me3 antibodies. Polyclonal antibodies raised against tri-methyl K27 showed specific reactivity with human (HsH3K27me3: SKAARK(Me3)SAP) or *Paramecium* tri-methyl K27 (PtH3K27me3: TKAARK(Me3)TAP) but not against no-methyl-, mono-methyl, di-methyl *Paramecium* K27, di-methyl, tri-methyl *Paramecium* K9 (QTARK(Me3)STAGN). **B.** Competition assay. 0 to 100 pmol of *Paramecium* H3K27me3 peptides were spotted and probed with the H3K27me3 antibodies alone or in presence of a 50-fold molar excess of the indicated peptides. Competition with the *Paramecium* (PtK27me3: TKAARK(Me3)TAP) or human (HsK27me3: SKAARK(Me3)SAP) tri-methyl K27 peptides completely eliminates the signal, while competition with the un-methylated peptide (PtK27me0: TKAARKTAP) does not.

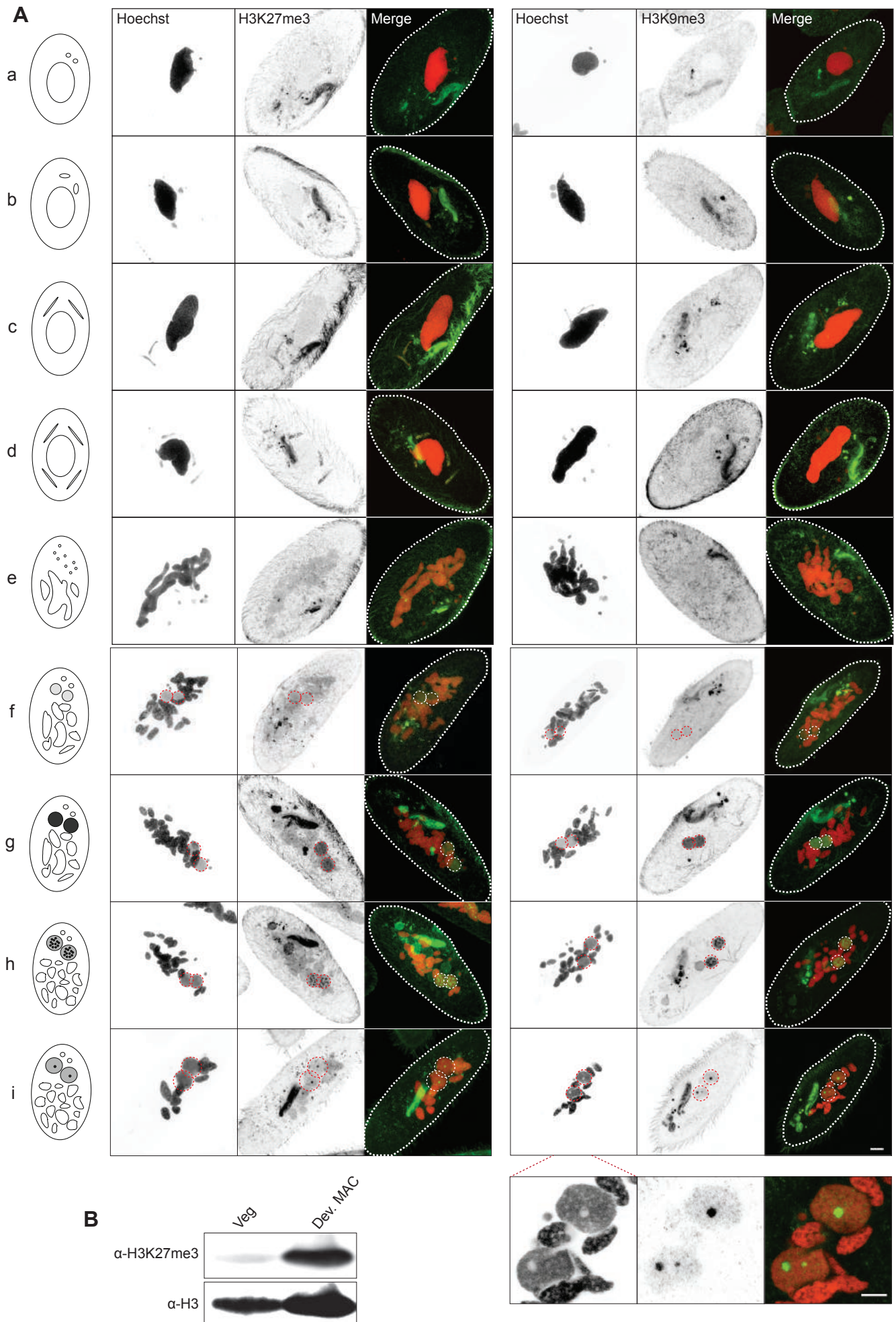


Figure S2. Immunostaining with H3K9me3 or H3K27me3 antibodies during *Paramecium* life cycle.

Schematic representations of key nuclear events in *Paramecium* autogamy are depicted on the left: (a) vegetative growth, (b-c-d) meiosis I, (e) meiosis II, (f-i) MAC development. **A.** Immunolabeling with H3K27me3 or H3K9me3 antibodies (in green) and staining with Hoechst (in red). Panels (f-i) are the entire images of the magnified views presented in Figure 1 (a-d). Scale bar is 10 μ m. Insets present magnified views of the specified areas and scale bar is 5 μ m. **B.** Western blot with H3K27me3 and H3 antibodies on acid extracts from vegetative nuclei (Veg) or developing somatic macronuclei (Dev. MAC) at T25 hrs during autogamy of wild type cells.

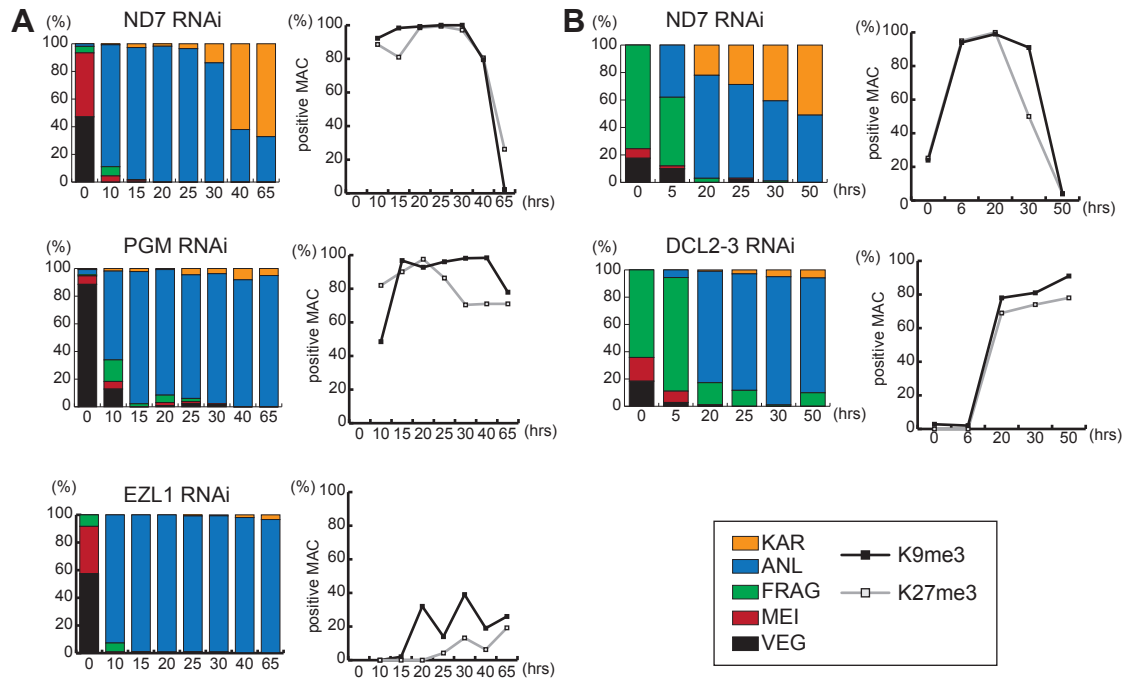


Figure S3. Progression of autogamy. Progression of autogamy was followed by cytology with Hoechst staining in independent time course experiments after **A.** ND7 (control), or PiggyMac (PGM) or EZL1 silencing and **B.** ND7 (control) or DCL2 and DCL3 silencing. Linear charts show quantifications of positive signals in the developing MAC after immunolabeling with H3K27me3 (grey line) or H3K9me3 (black line) antibodies at each time point. T0 hr corresponds to 50% of cells with fragmented old MAC and the time-points refer to hours after T0. VEG: vegetative, MEI: meiosis, FRAG: fragmented maternal MAC, ANL: two visible developing MAC, KAR: Karyonide. At least 100 cells were scored for each time point by fluorescence microscopy.

Figure S4

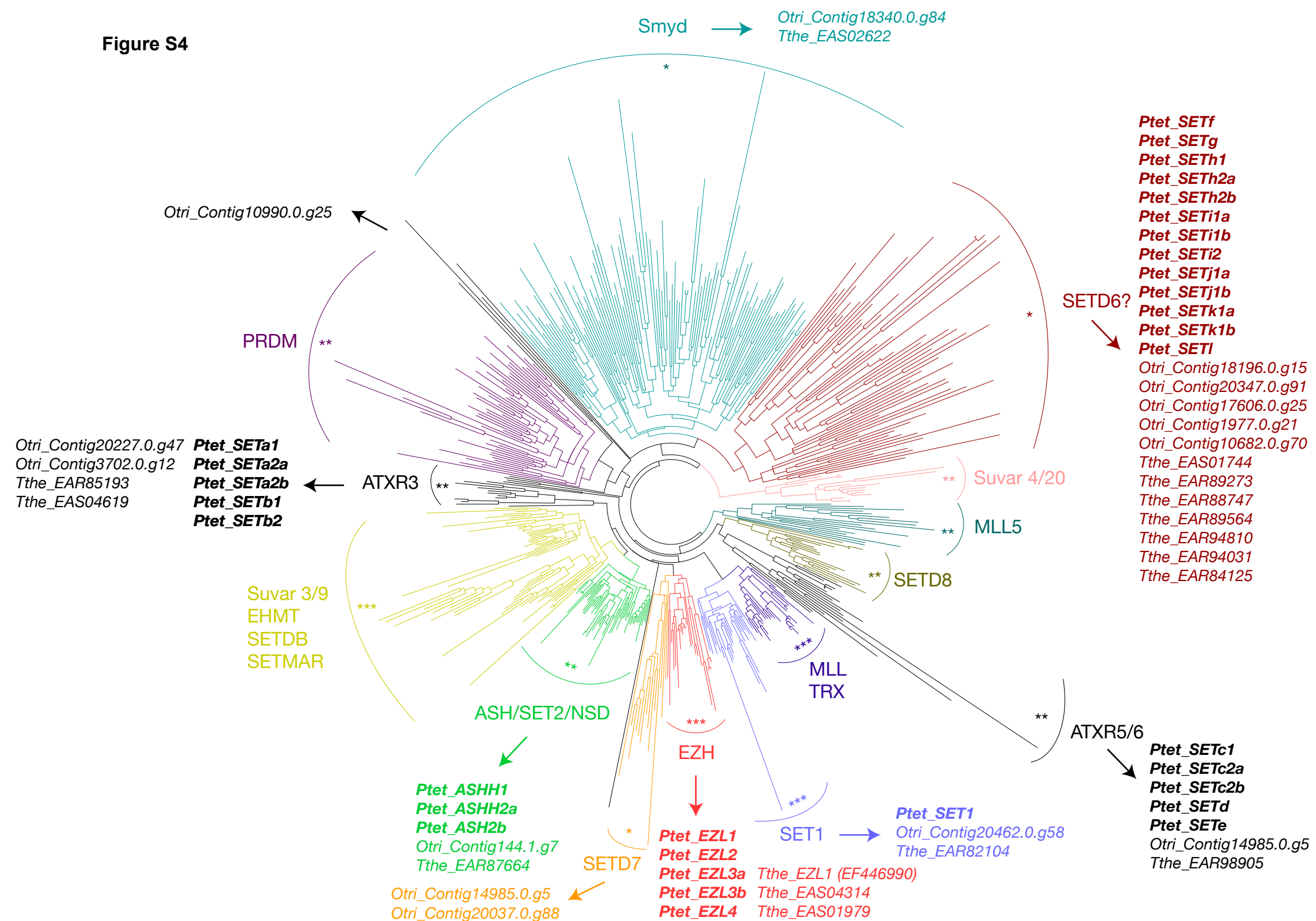


Figure S4. Phylogenetic analysis of SET domain proteins from *P. tetraurelia*, *T. thermophila* and *O. trifallax*. An unrooted Maximum-likelihood (ML) tree is shown. This tree has been constructed using the SET domains encoded by the genome of a large number of species. The full list of used sequences and studied species can be found in Table 1 and Fig. S13. We identified 13 monophyletic groups that include sequences from several different species. We named most of these groups by using the name(s) of the Human and/or *Drosophila* proteins that are included in the group. We tried to use as much as possible the nomenclatures used in previous phylogenetic analyses of SET domain proteins (Sun et al., 2008; Zhang and Ma, 2012). In two cases, the monophyletic groups do not include animal proteins and we used the name(s) of the included *Arabidopsis* protein(s) to name these groups (ATXR3 and ATXR5/6). We listed in the figure the proteins from *Paramecium tetraurelia* (*Ptet*), *Oxytricha trifallax* (*Otri*), and *Tetrahymena thermophila* (*Tthe*), which belong to the different monophyletic groups. The robustness of the nodes that define the different monophyletic groups was assessed by evaluating their statistical support (aLRT values) in the ML analysis and by performing phylogenetic analyses (ML and Bayesian inference) using a smaller sampling of species (ciliates + Human + *Drosophila* + yeasts; not shown). This is represented on the tree by the presence close to the name of the group of *** (aLRT values > 0,8 and groups similar in all analyses), ** (0,8 > aLRT values > 0,5 and groups similar in all analyses), or * (aLRT values < 0,5 and/or groups significantly different in the analyses with a different sampling of species).

We obtained strong support for the existence of ciliate members of the EZH, ASH/SET2/NSD and SET1 groups that also include proteins from several other species including animals and *Arabidopsis*. Several ciliate SET domains cluster with either *Arabidopsis* ATXR3 or *Arabidopsis* ATXR5/6 in groups that only include sequences from a small number of species and none from Human or *Drosophila*. These groups may therefore correspond to ancestral SET domain proteins that have been lost in some lineages such as animals or to divergent members of some other groups. This latter possibility is supported by the fact that the two groups are associated to the ASH/SET2/NSD group in the phylogenetic analyses made on the ciliates + Human + *Drosophila* + yeasts dataset (not shown). The inclusion of *T. thermophila* and *O. trifallax* proteins in the SMYD and SETD7 groups, as well as the belonging of a large number of ciliate proteins to the SETD6 group, has to be taken with caution, given the poor support of these groups, and may correspond to the artefactual grouping of highly divergent sequences.

[illegible]

Tthe_EAS01979	20	-----LNANINVTIRSHLTKS	KVVIANS	ETCGG	-LGIYINVYPIQKGDLLITY
Tthe_EAS04314	19	-----LNKNPNLIFYFMLSK---	RTVIGMSFVCQG	-LGLYNVFPQQHSLVLSY	
Ptet_EZL4	39	-----ICSNTQILIKNVK----	PTLLARSTVCSC	-LGLFSKHFIMKCELTILY	
Ptet_EZL3a	39	-----VCSNTQILINNVK----	PTLLGRSGVCNG	-IGVFARNYIMKDELILY	
Ptet_EZL3b	41	-----VCSNTQILINNVQ----	PTLLGRSGVCNG	-LGVFARNYIMKDELILY	
Ptet_EZL2	45	-----NCNNVNLTLGISK----	QLTLGNSLTCNG	-IGLFAAQHFKTYDFIGEY	
Ptet_EZL1	46	-----SCSNKNLLMQIRK----	SLVLGKSLTCNG	-LGLFAAQNFVKCDFVGEY	
Ttra_AMSG_09360T0	42	-----PCRNLAHFNVKVN----	KVVVGISTVAG	--WGLFAGQPISEKQFIGEY	
Tthe_EZL1	27	IQKFGLNID---	YCTNSQALYNCKP	--RVLLGKSLVCEG	-LGLFAGQDFKKNQYIGCY
Otri_Contig14971_0_g38	20	-----LCKNPLTLNFPCK----	RVFTAKSTLCDEIV	GLFTMDRVKDELIMEY	
Amac_AMAG_02871T0	20	-----ACKNREIGKGVHDHKKLR	VDVSTIPKAG	--WGLFAQCDFKADDFIGEY	
Spun_SPPG_03589T0	57	AVGRKTSSL---	TCHNQAMQRNMKK----	HLVCGDSVHVG	-RGAFLREPAEKCELAIEY
Atha_MEA	35	TLGETPVQI---	QCKNQFLLQTNK----	KILLGKSDVHVG	-WGAFTWDSLKKNEYLGEY
Atha_CLF	27	SLGVPSQRGDNYE	CRNKKLLKQQQ----	RVLGLISDVSG	-WGAFLKNSVSKHEYLGEY
Atha_SWN	31	SLGEAPRRGEG	QCGNMLLLRQQQ----	RILLGKSDVAG	-WGAFLKNSVSKNEYLGEY
Mbre_estExt_fgenesH2_pg.C_20049	44	-----VCVCVCVCVCVCVC	VCVCVCVCVCVC	-----VCVCVCVCVCVCVC	
Dmel_Ez	51	TKI-----	TCKNVQVQRLHKK	-HLLMAPSDIAG	-WGIFLKEGAOKNEFISEY
Lgic_estExt_Genewise1.C_sca_1470059	34	IKI-----	SCKNVSVQRSQCK	-HLLLAPSDVAG	-WGIFLKEPAEKNEFISEY
Nvec_estExt_GenewiseH_1.C_5680002	49	DSK-----	TCKNVSLQRQQRK	-HLLLAPSDVAG	-WGIYIKQSFKNEFISEY
Aque_Aqu1_209566	38	MKFT-----	TCKNVSIQRGQKK	-HLLMVLSDVAG	-WGIFLKDGAEKNEFISEY
Hsap_EZH1	53	CKVV-----	SCKNCSIQRGLKK	-HLLLAPSDVAG	-WGTFIKESVOKNEFISEY
Hsap_EZH2	51	SKNV-----	SCKNCSIQRGSKK	-HLLLAPSDVAG	-WGIFIKDPVOKNEFISEY

AdoMet G
XG
catalytic Y

```

Tthe_EAS01979      367 YGEVLQDLDIIR---DSWKPSNLFYIFSLLG-DL--TVDSKYIGNKSRFMNHSKS---K
Tthe_EAS04314      63 LGEYISDEQVIR---ESWLKSDIFYTFTKNDFEK--LVDSKYFGNKSFRFMNHHISLE--
Ptet_EZL4          82 IGEAIIIDDEDIIR---DQFDDAFSFFNYQLSDERY--SLDSRFCGNESRFINHNSL--NL
Ptet_EZL3a         82 ICEVLIDDEDIIR---DQFDDTFTFFYNLSLNDKY--SLESRFCGNESRFINHNSS--NL
Ptet_EZL3b         84 ICEVLIDDEDIIR---DQFDDTFTFFYNQLNEDKY--SLESRFCGNESRFINHNSQ--NL
Ptet_EZL2          88 RCNYLLLEDSESYI-IEQCNDLTGKHYLFEVDD-KW--QVDGTYYSNYLRFINHATNQSET
Ptet_EZL1          89 TGNYLILLDDDESMA-IEQCDWITNNHYLFEVDD-KW--QVDGTYYSNCLRYINHATKKSDDL
Ttra_AMSG_09360T0  84 VGEVVSSTVEERR--HSEYSSMRISYVFDYNR-DF--VIDAFRKGSGTKLIHNSDE----
Tthe_EZL1          79 IGEIINEKQGTFR--QBWQQPGQGISYLEMLNK-ET--DVDSFRFYGNKMRVYVNHCG--SM
Otri_Contig14971_0_g38 65 TGKVIITQEGQKETVDQLVNDLGRGRSYGFTLDR-LT--TLDAVYVGNLMRFANHS SD--KL
Amac_AMAG_02871T0  66 SGEVLSDEADRR--GIHYDRKSLNYCFOLAT-DA--VVDAIRLGPVLRFCNHA AK----
Spun_SPPG_03589T0  108 TGGVIVSAESNRR--DAITNAASKLNYLFSLGD-KYAPLVDA SRKGNKIRFINHGEY-PDK
Atha_MEA          86 TCELIHDEANER--GRIEDRIGSSYLFTLND-QL--EIDARRKGNEFKFLNHSAR----
Atha_CLF          81 TCELI SHKEADKR--GKIYDRENCSEFLFNLD-QF--VLDAIRKGDKLKFANHSPE----
Atha_SWN          84 TCELI SHHEADKR--GKIYDRANSSFLFDLND-QY--VLDAORKGDKLKFANHSAK----
Mbren_estExt_fgenes h2_pg.C_20049 73 CVCIIISQEEADRR--GKVYDQLKCSFLFNLNQ-EY--VVDAIRKGNKIRFANHAND----
Dmel_Ez          96 CGEIIISQDEADRR--GKVYDKYMCSEFLFNLN--DF--VVDAIRKGNKIRFANHSIN----
Lgic_estExt_Genewis e1.C_sca_1470059 79 CGEIIISQDEADRR--GKVYDKYMCSEFLFNLN--DF--VVDAIRKGNKIRFANHSIN----
Nvec_estExt_Genewis eH_1.C_5680002 94 CGEVISQDEADRR--GKVYDKYMCSEFLFNLN--DF--VVDAIRKGNKIRFANHSIS----
Aque_Aqu1_209566  84 CGEIIISQDEADRR--GKVYDKYMCSEFLFNLN--DY--VVDAIRKGNKIRFANHSVD----
Hsap_EZH1         99 CGELISQDEADRR--GKVYDKYMSSFLFNLN--DF--VVDAIRKGNKIRFANHSVN----
Hsap_EZH2         97 CGEIIISQDEADRR--GKVYDKYMCSEFLFNLN--DF--VVDAIRKGNKIRFANHSVN----

```

Tthe_EAS01979	118	ENSYAKMVYAKGGYTIGLFSNEKIIPGTELELFDYD-GQGTLKLKYEWIMQKADS---KAK
Tthe_EAS04314	116	-NCFARSVYSRGDYHGLYAKQDIPPCNELFDYD-ADKTLAVEKEWITRNNYSN--YKI
Ptet_EZL4	135	NNCRTNQIFTCGQYQLAIFAKNIDPEQETLLNNEGESLNKEVHNWNEQQQ-----QFW
Ptet_EZL3a	135	MNCKTRQIFSSGKFQLAIFYALKDIYPEQETLLNNEGDQINRDLNNWVDINT-----QYW
Ptet_EZL3b	137	MNCKTRQIFSSGKFQLAIFYALTEISPQQETLLNNEGDQINKDLNNWVDMNS-----QYW
Ptet_EZL2	144	ANCOAILLFSBGRWKIGMLATREIEVGQELVFDYG-----DKFKTKLQEFNKISE-RYF
Ptet_EZL1	145	ANCOAQILFSBGRWRIRAMFTTKNISTGEELFFDYG-----DKFLTTLTDFNKLCD-DYY
Ttra_AMSG_09360T0	135	PNVHPMTKVVNGVERVGLYASRDILRTGDELFFCYN----YDKEKREWYLDSDFQQILEAE
Tthe_EZL1	132	ANCKVDVITYNRGINIVRFSAKEDIQKQGETYFDYN-----RNYQIEWMKYFNEYD-KFE
Otri_Contig14971_0_g38	120	SNCKINMIFAQGMQRVCLIVASRNIDQGEELFFDYG-----FSKEFEWLKDYDRNGCIDVI
Amac_AMAG_02871T0	117	SNVNAHVAFVDSQSRIGLYAKKNINAGDELFDYGRMYWGDESGETPKAAASTSSSSSSPS
Spun_SPPG_03589T0	164	ANCFPVVKMVSGELRVGIYARRIEAEKELFFDYG-----EDF
Atha_MEA	137	PNCYAKLMIVRGDQRIGLFABRAIEEGEELFFDYC--YGPEHADWSRGREPRKTGASKRS
Atha_CLF	132	PNCYAKVMIVAGDHRVGIFAKERILAGEELFFDYR--YEPDRAP-AWAKKPEAPGSKKDE
Atha_SWN	135	PNCYAKVMFVAGDHRVGIFAKERIEASEELFFDYR--YGPDAQP-VWARKPE--GS-KKD
Mbre_estExt_fgenesh2_pg.C_20049	124	PNCCARVMMVAGEHRIGIFABRDIPACRELFNRYR--YGPTDAL-----KYV
Dmel_Ez	147	PNCYAKVMVVGCDHRIGIFAKRAIQGEELFFDYR--YGPTQL-----KFV
Lgig_estExt_Genewise1.C_sca_1470059	130	PNCYAKVMVNGDHRIGIFAKRSIQSCEELFFDYR--YGPTQL-----RFV
Nvec_estExt_GenewiseH_1.C_5680002	145	PNCYAKVMVNGDHRIGIFAKRDIEAGEELFFDYR--YSATDAL-----KFV
Aque_Aqu1_209566	135	PNCYAKVMIVGCDHRIGIFAKHNIIELEELFFDYR-----
Hsap_EZH1	150	PNCYAKVMVNGDHRIGIFAKRAIQAGEELFFDYR--YSQADAL-----KYV
Hsap_EZH2	148	PNCYAKVMVNGDHRIGIFAKRAIQTCEELFFDYR--YSQADAL-----KYV
		lysine binding EELXFDY
Tthe_EAS01979	174	QIQEDNIIVFEAHNSPAKNQNLKQKKRK---
Tthe_EAS04314	172	RMDNIRITIPRNKDLTRFNSTLIIDEN---
Ptet_EZL4	190	NYIQNTDTKER-----
Ptet_EZL3a	190	NYRQTTLIGKK-----
Ptet_EZL3b	192	NYKQTTFIG-----
Ptet_EZL2	198	QGK-----
Ptet_EZL1	199	KK-----
Ttra_AMSG_09360T0	191	GNKEAPLVVE-----
Tthe_EZL1	186	QEEKKQQVQKKIISKFIKN-----
Otri_Contig14971_0_g38	175	VIKQPDGTLVSSPFHLRFGLKVLKSSDK---
Amac_AMAG_02871T0	177	SPAAASSRRKRARK-----
Spun_SPPG_03589T0	202	TK-----
Atha_MEA	195	KEARPAR-----
Atha_CLF	189	NVTPSVGRPKKLA-----
Atha_SWN	189	DSAITHRRARKHQSH-----
Mbre_estExt_fgenesh2_pg.C_20049	169	SVERDQEEEPVDWLDIDWVQPLSDFTGVTMAV
Dmel_Ez	192	GIEREMEIV-----
Lgig_estExt_Genewise1.C_sca_1470059	175	GIERDADVP-----
Nvec_estExt_GenewiseH_1.C_5680002	190	GIERDVDFALR-----
Aque_Aqu1_209566		-----
Hsap_EZH1	195	GIERETDVL-----
Hsap_EZH2	193	GIEREMEIP-----

Figure S5: Alignment of SET domains from EZH-EZL proteins from *P. tetraurelia*, *T. thermophila*, *O. trifallax* and other organisms. Sequences were aligned using the Muscle v3.8 software. The white text on a black background denotes invariant residues; black text on a gray background indicates conserved residues. Highly conserved residues are highlighted with different colors according to Dillon, 2005 and Joshi et al. 2008: catalytic site (red), adenosylmethionine (AdoMet) binding pocket (green), lysine substrate binding pocket (blue). Note that Ptet_EZL3a, Ptet_EZL3b and Ptet_EZL4 do not show all conserved residues. Accession numbers are given in Table S1 and Fig. S. Species name abbreviations: Amac = *Allomyces macrogynus* (Fungi); Atha = *Arabidopsis thaliana* (Viridiplanta); Aque - *Amphimedon queenslandica* (Metazoa); Dmel = *Drosophila melanogaster* (Metazoa); Hsap = *Homo sapiens* (Metazoa); Lgig = *Lottia gigantea* (Metazoa); Mbre = *Monosiga brevicolis* (Choanoflagellata); Nvec = *Nematostella vectensis* (Metazoa); Otri = *Oxytricha trifallax* Ptet = *Paramecium tetraurelia* (Ciliata); Spun = *Spizellomyces punctatus* (Fungi); *Tetrahymena thermophila* (Ciliata); Ttra = *Thecamonas trahens* (Apusozoa)

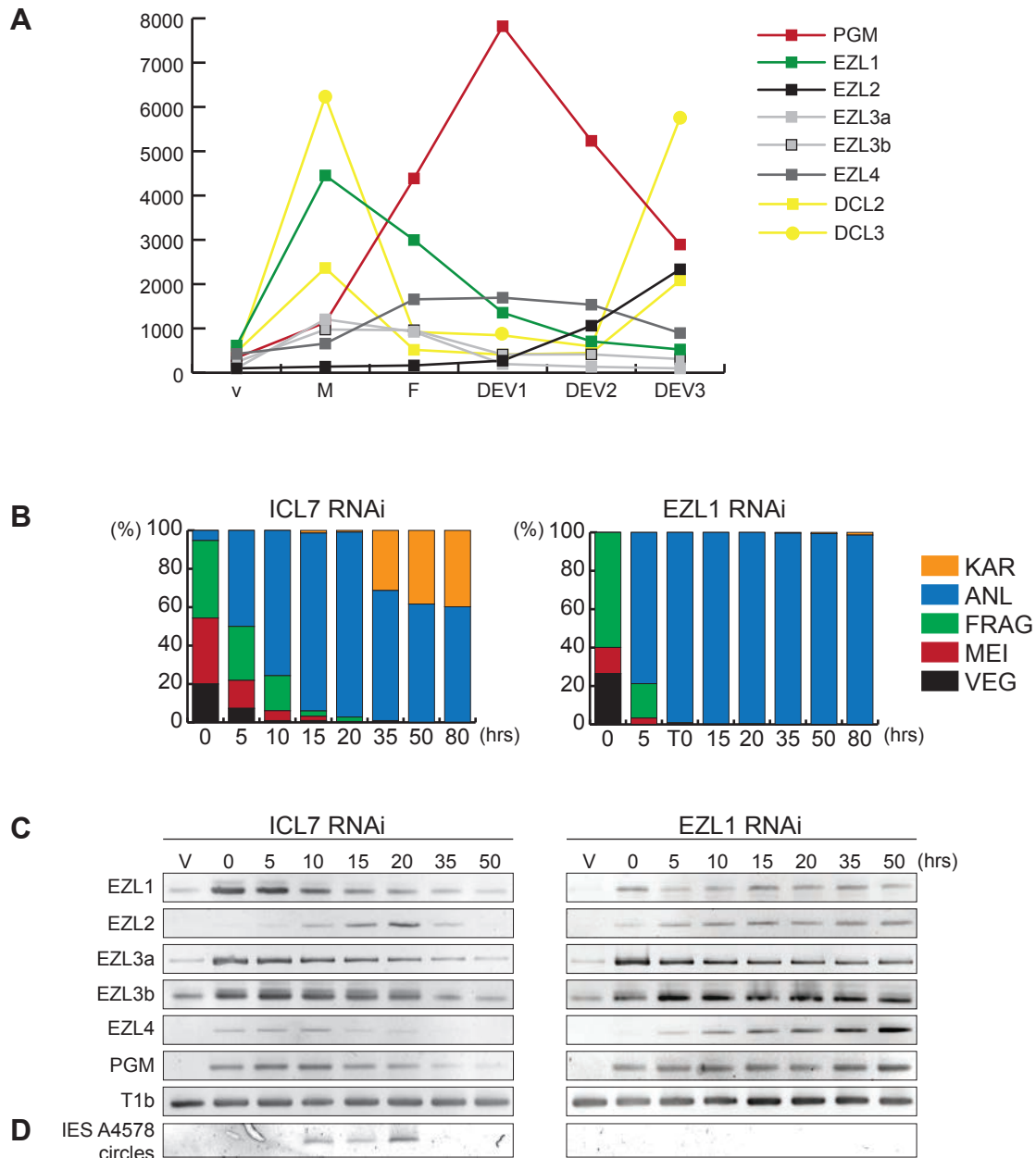


Figure S6. Expression patterns of EZL genes during the life cycle.

A. EZL, DCL2, DCL3 and PGM gene expression levels, as determined by microarray expression data during autogamy time-course experiments (Arnaiz et al., 2010). The vegetative time point (VEG) consists of 4 samples from mass cultures containing only log-phase cells showing no sign of meiosis. The meiosis time point (MEI) consists of 4 samples containing 20-39% of cells undergoing meiosis, and little or no fragmentation of the old MAC. The fragmentation (FRAG) time point consists of 4 samples that contained a similar proportion of meiotic cells (20-29%) as the MEI time point, but also contained 37-43% of cells with a fragmented old MAC. The DEV1 time point groups 3 samples with 35-56% of cells with fragmented old MACs and 35-51% of cells that already contained clearly visible new MACs (anlagen). DEV2 consists of 3 samples with 73-98% of cells with visible anlagen, and the DEV3 samples were taken ~10 hours after the DEV2 samples. **B.** Progression of autogamy in two independent time course experiments, in which ICL7 (control) or EZL1 genes have been knocked down by RNAi. See Figure S3 legend. **C.** Detection of EZL and PGM mRNA during autogamy by RT-PCR. Total RNAs were extracted at each time point shown in (B), were reverse transcribed and cDNAs were amplified by PCR with gene specific primers and, as a loading control, with primers for the T1b gene, which encodes a component of the secretory granules (Madeddu et al. 1995). **D.** PCR detection of IES 51A4578 circles with divergent primers on genomic DNA at each time point shown in (B) after ICL7 (control) or EZL1 silencing.

Figure S7

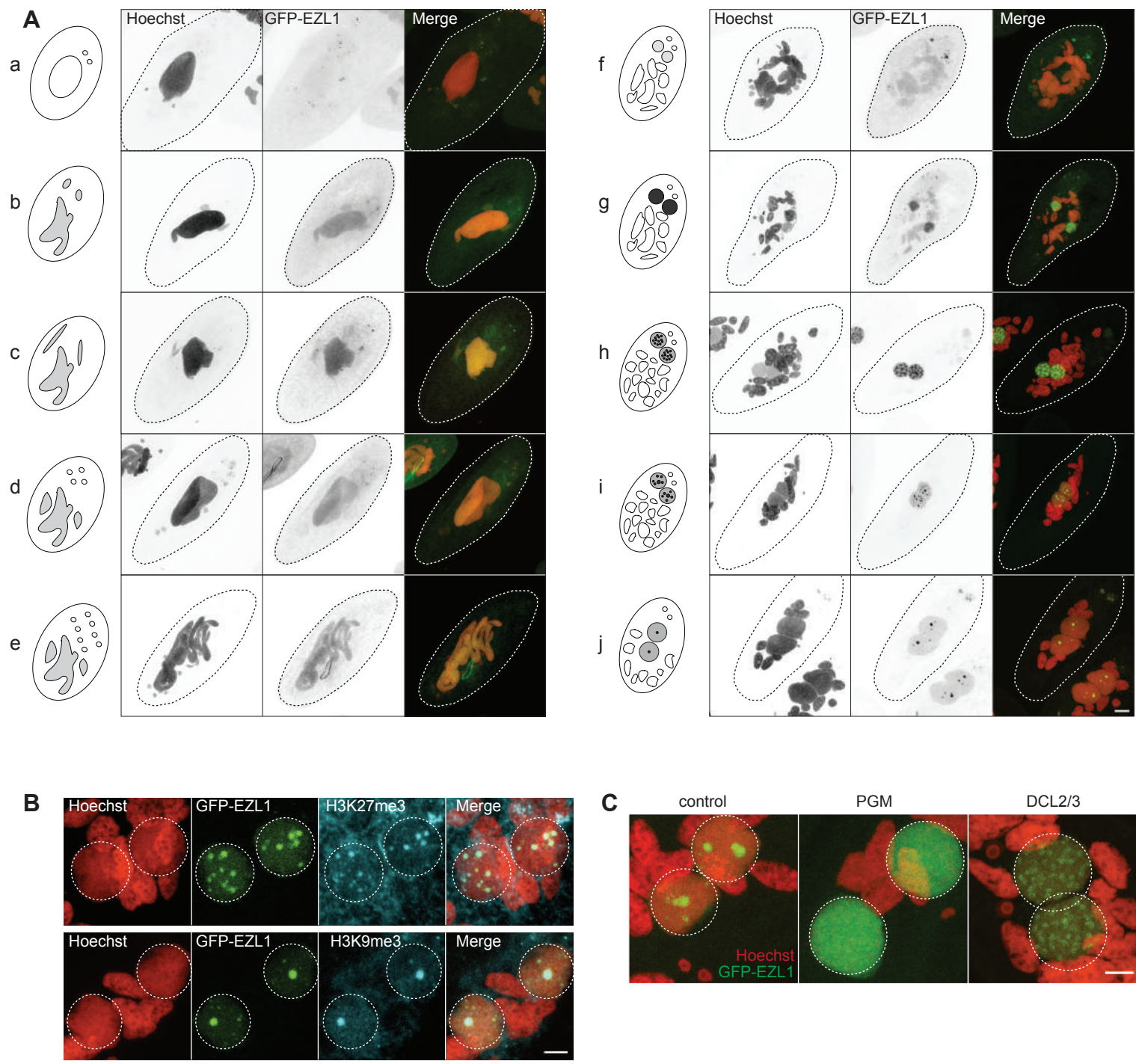


Figure S7. Localization of a GFP-EZL1 fusion protein. **A.** Localization of a GFP-EZL1 fusion protein during vegetative growth (a), meiosis I and II (b-e) and MAC development (f-j). Panels (f-i) are the entire images of the magnified views presented in Figure 3 (a-e). Scale bar is 10 μ m. **B.** Co-localization of H3K27me3 and H3K9me3 with the GFP-EZL1 fusion protein. GFP-EZL1 transformed cells were immunolabeled with H3K27me3 or H3K9me3 antibodies and stained with Hoechst at T40 hrs during autogamy. Overlay of Z-projections of magnified views of GFP-EZL1 (in green), H3K27me3- or H3K9me3- specific antibodies (in blue), and Hoechst (in red) on selected stacks are presented. Dashed white circles indicate the two developing MAC. The other Hoechst-stained nuclei are fragments from the old vegetative MAC. Scale bar is 5 μ m. **C.** GFP-EZL1 fusion protein localization after PGM or DCL2/3 silencing at T40 hrs during autogamy. Scale bar is 5 μ m.

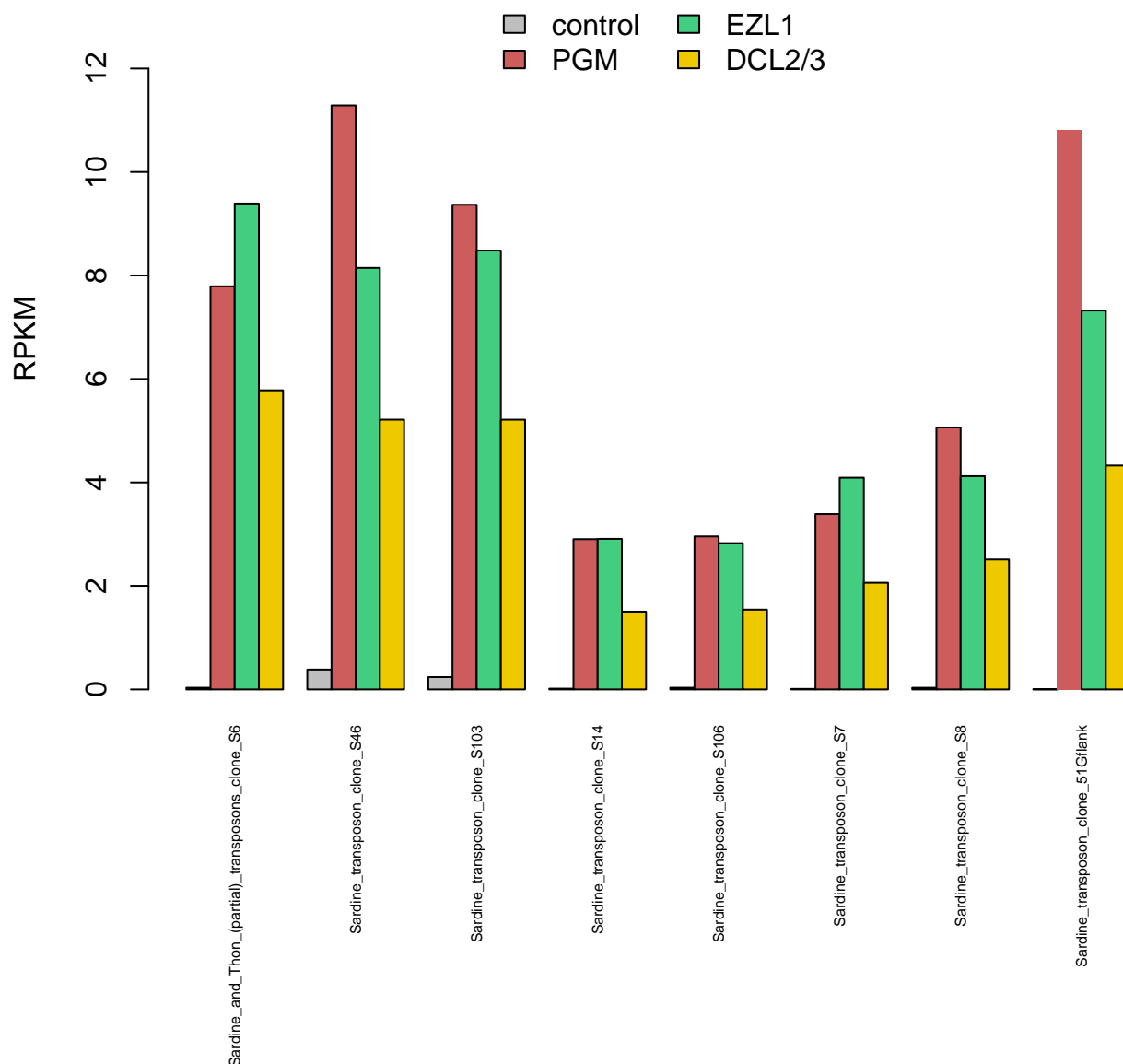


Figure S8. Sardine coverage.

The bar plots represent read coverage of 8 individual copies of the Sardine transposon (GenBank Accession No. HE774468-HE774475). The coverage was determined by mapping reads using BWA with default parameters, for the control dataset (grey), PGM silencing (red), EZL1 silencing (green) and DCL2/3 co-silencing (yellow) datasets. The normalized units (RPKM) are reads per kilobase of the transposon sequence per million library read mapped against the MAC reference genome.

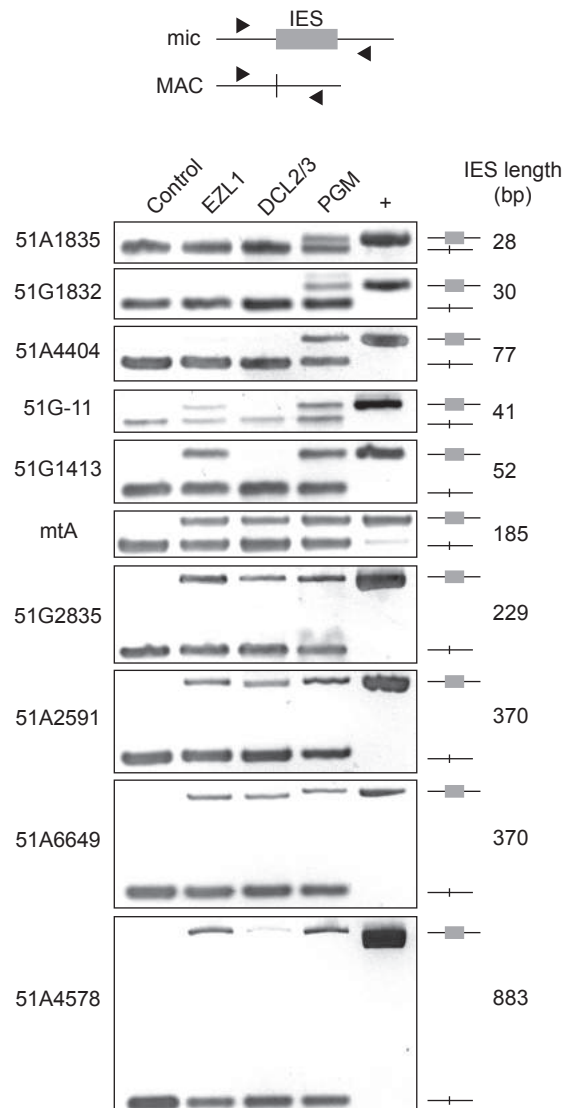


Figure S9. PCR analysis of IES retention after EZL1, DCL2/3 or PGM silencing.

PCR analysis of IES retention with primers (black arrows, Table S7) located on either side of the IES in mass autogamies after RNAi-mediated silencing of the indicated genes. Total DNA samples were prepared from starved post-autogamous cells at approximately T72 hrs. Because the parental MAC is still present at this stage, the excised version is amplified in all cases; the IES-retaining fragment can be detected only if it accumulates in the zygotic MACs. Control: unrelated negative control (ND7 or ICL7 RNAi); “+”: positive PCR control on cloned MIC DNA, or DNA from mating-type E cells in the case of the mtA IES.

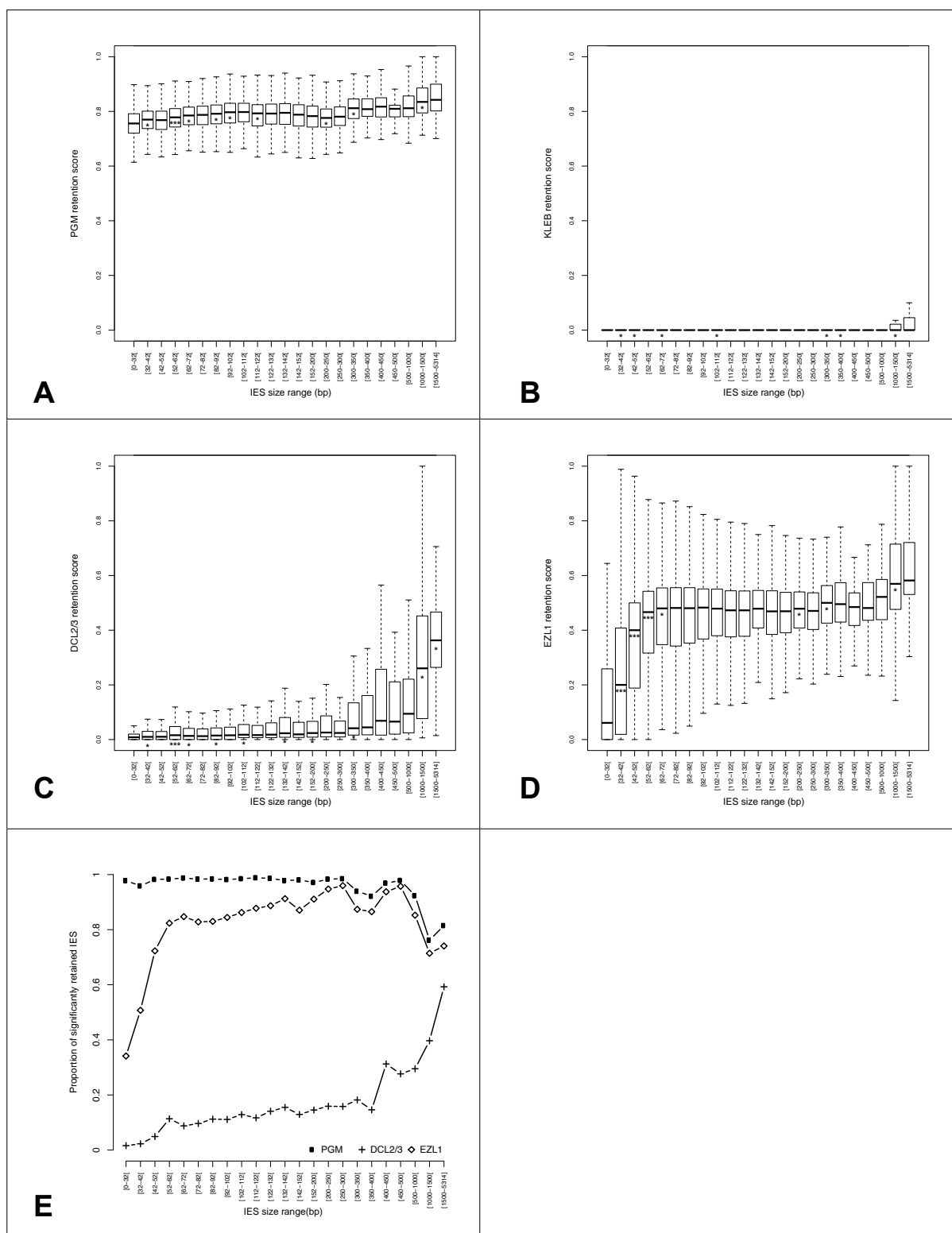


Figure S10. IES retention score and size in control, EZL1, DCL2/3 or PGM datasets.

IESs were grouped by size (cf. legend of Figure 5B), and boxplots were determined to show the distribution of **A**. PGM dataset retention scores, **B**. control dataset retention scores, **C**. DCL2/3 dataset retention scores and **D**. EZL1 silencing sample retention scores (reproduced from Fig. 5B to facilitate comparison). **E**. The proportion of significantly retained IESs for PGM, DCL2/3 and EZL1 samples are plotted as a function of IES size, using the same size groups as for parts A-D, corresponding to the peaks in the IES size distribution. This representation shows, for example, that 34 % of the IESs in the smallest peak are significantly retained after EZL1 silencing. In the PGM dataset, essentially all IESs are significantly retained except in the the last few groups of largest size, and that is because of the small number of IESs in those groups, introducing a lot of noise in the curve. In the EZL1 dataset, the increase in significantly retained IESs is dramatic over the first 3 groups in the size distribution and then flattens out. The variation for the largest size groups is very similar to the variation found for the PGM dataset, in support of the hypothesis that this variation is noise that arises from the small number of IESs in those groups. For the DCL2/3 dataset, few IESs are significantly retained in any of the groups. However, there does appear to be an increase in significantly retained IESs for the largest IES sizes. As already noted, the largest groups include very few IESs introducing a lot of noise, however for the DCL2/3 dataset, the variations are different than for the other 2 samples, in support of a real increase in significant retention of IESs larger than ~ 500 bp. In A-D, the stars indicate that the retention score distribution of a group is significantly different from the distribution of the previous group.

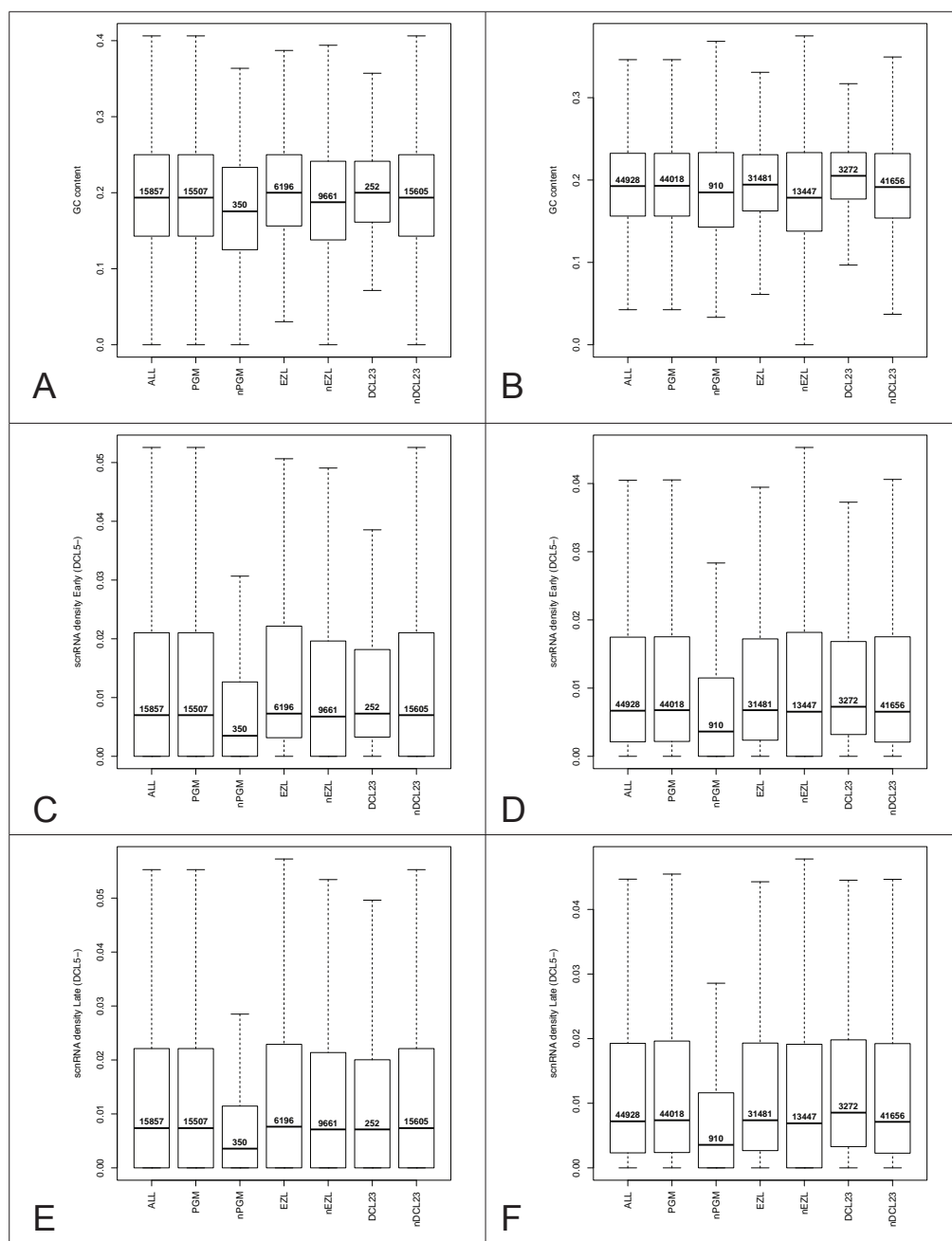


Figure S11. IES properties are compared using boxplots of IESs grouped according to sample. The left column (A,C,E,G) concerns only the IESs in the first peak of the size distribution (< 32 bp) and the right column (B,D,F,H) considers all IESs whatever their size. The IESs in each group are those with significant retention scores in the sample under consideration (PGM, EZL, DCL2/3), IESs that do NOT have significant retention scores in the same samples (nPGM, nEYL, nDCL2/3) or “ALL” IESs. These IES groups were examined with respect to different IES properties: GC content (A,B), the density of 25 scnRNAs from an early time point (cf. Sandoval et al., 2014) that map to the IESs (C,D), the density of 25 nt scnRNAs from a late time point (cf. Sandoval et al., 2014) that map to the IESs (E,F). The number of IESs in each group is indicated in boldface type inside the boxplots. See Sup. Materials and Methods for more information about the small RNA samples and data analysis.

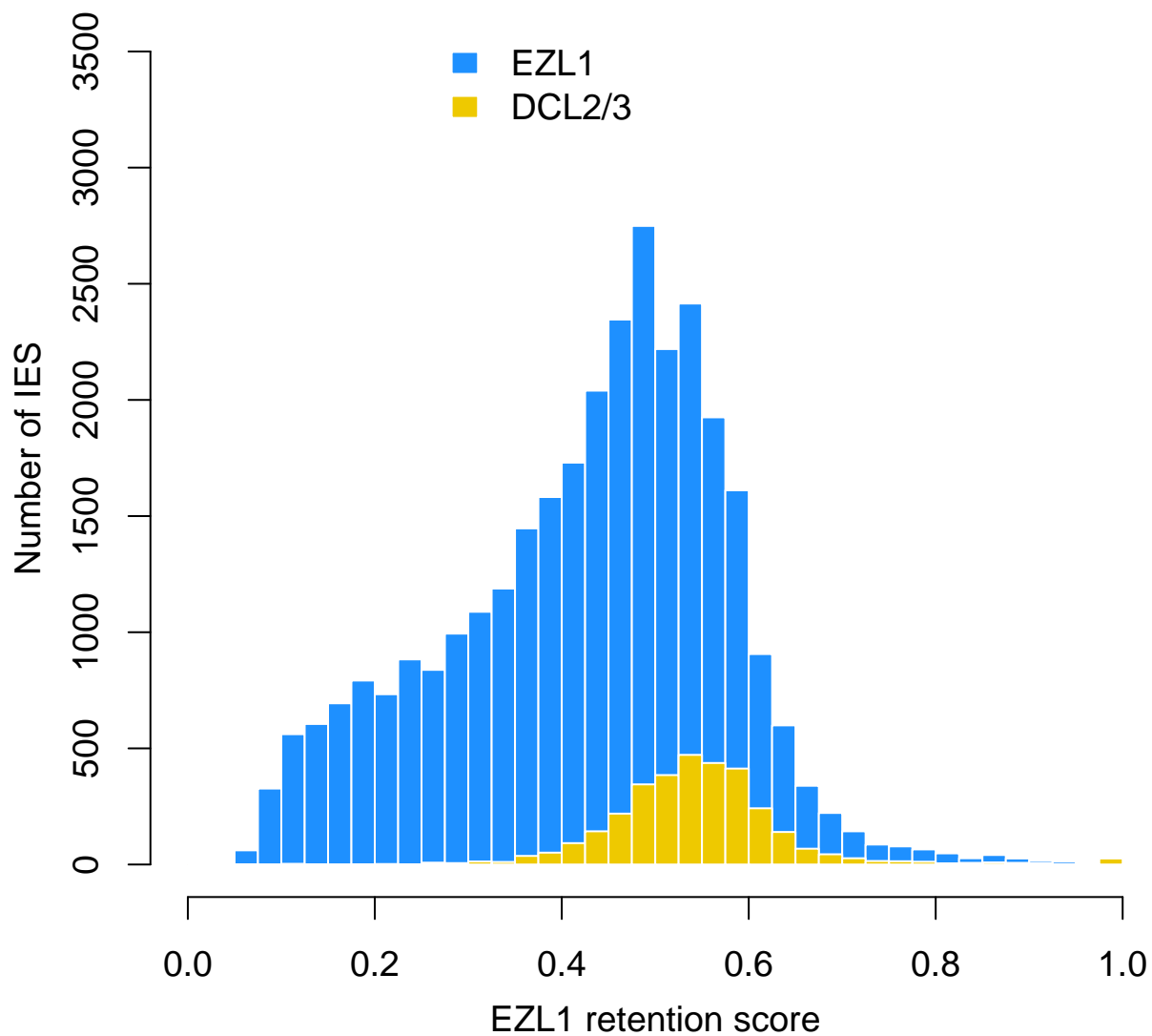


Figure S12. EZL1 retention score for DCL2/3 retained IESs.

Stacked histogram of EZL1 retention scores for EZL1-retained IESs (blue) and DCL2/3-retained IESs (yellow), showing that the significantly retained IESs in the DCL2/3 dataset are among the most significantly retained IESs in the EZL1 dataset.

Paramecium SET domain proteins	Paramecium DB Accession numbers	Domains (CDD-search + SMART)
Ptet_EZL1	GSPATG00037872001	SET: 459-571
Ptet_EZL2	GSPATG00032888001	SET: 466-575
Ptet_EZL3a	GSPATG00012695001	Metallothio: 467-509 SET: 555-661
Ptet_EZL3b	GSPATG00013305001	Metallothio: 460-509 SET: 555-661
Ptet_EZL4	PTETG1700020001	SET: 553-659
Ptet_SET1	GSPATG00013040001	zf-HC5HC2H: 347-417 SET: 704-821
Ptet_ASHH1	GSPATG00013040001	AWS: 28-62 SET: 73-187 PHD: 223-271
Ptet_ASHH2a	GSPATG00003275001	AWS: 67-127 SET: 138-252 PHD: 288-336
Ptet_ASHH2b	GSPATG00004957001	AWS: 67-127 SET: 138-252 PHD: 288-336
Ptet_SETa1	GSPATG00014017001	SET: 283-399
Ptet_SETa2a	GSPATG00025368001	SET: 120-242
Ptet_SETa2b	GSPATG00031547001	SET: 306-414
Ptet_SETb1	GSPATG00018768001	Bromodomain: 133-183 SET: 422-544
Ptet_SETb2	GSPATG00035094001	Bromodomain: 138-190 SET: 421-553
Ptet_SETc1	GSPATG00025951001	PHD: 40-80 SET: 196-308
Ptet_SETc2a	GSPATG00033097001	PHD: 41-73 SET: 245-357
Ptet_SETc2b	GSPATG00036078001	PHD: 36-82 SET: 254-366
Ptet_SETd	GSPATG00035182001	SET: 296-402
Ptet_SETe	GSPATG00021145001	SET: 19-135
Ptet_SETf	GSPATG00024000001	TPR: 142-238 SET: 303-466
Ptet_SETg	GSPATG00031631001	SET: 156-230
Ptet_SETh1	GSPATG00003961001	SET: 175-270 Rubis-subsb-ind: 299-429
Ptet_SETh2a	GSPATG00027281001	SET: 182-270 Rubis-subsb-ind: 299-429
Ptet_SETh2b	GSPATG00020579001	SET: 175-270 Rubis-subsb-ind: 299-429
Ptet_SETi1a	GSPATG00015535001	TPR: 13-111 SET: 153-359 Rubis-subsb-ind: 395-531
Ptet_SETi1b	GSPATG00012103001	TPR: 13-111 SET: 153-369 Rubis-subsb-ind: 393-522
Ptet_SETi2	GSPATG00016863001	SET: 342-388 Rubis-subsb-ind: 423-545

Ptet_SETj1a	GSPATG00037631001	TPR: 161-194 SET: 251-383
Ptet_SETj1b	GSPATG00010387001	SET: 249-402
Ptet_SETk1a	GSPATG00038239001	SET: 34-253
Ptet_SETk1b	GSPATP00007235001	SET: 34-133
Ptet_SETk2a	GSPATG00012599001	SET: 34-527 Rubis-subs-bind: 610-692
Ptet_SETk2b	GSPATG00013397001	SET: 45-107
Ptet_SETI	GSPATG00012308001	SET: 34-450 Rubis-subs-bind: 526-609

Table S1. SET domain proteins identified in the somatic MAC genome of *P. tetraurelia*. Accession numbers (see ParameciumDB, <http://paramecium.cgm.cnrs-gif.fr/>), names and conserved domains are indicated.

Target gene	None	control	EZL1		EZL2	EZL3a	EZL3b	EZL3a+ EZL3b	EZL4
Gene fragment			#1	#2					
Viable (%)	97,5	97	3	0	97	94	99	96	96
Total nb of cells	480	534	717	120	108	129	108	48	108
Nb of experiments	10	11	13	2	2	2	2	1	2

Table S2. Survival of post-autogamous progeny following EZL gene silencing. The gene targeted in each silencing experiment is indicated. Two non-overlapping fragments (#1 and #2) of EZL1 gene were used independently and similar results were obtained with both constructs. The ND7 or ICL7 genes were used as control RNAi targets, since their silencing has no effect on sexual processes (Bouhouche et al., 2011). Autogamy was also performed in standard *K. pneumoniae* medium (none). Cells were starved in each medium to induce autogamy and, following 3–4 days of starvation, autogamous cells were transferred individually to *K. pneumoniae* medium to monitor growth of sexual progeny. The table summarizes the data from several independent experiments as indicated and shows the total number of autogamous cells analyzed for each RNAi. Death in progeny after EZL1 silencing was observed after less than three cell divisions.

DNA	Insert size (bp)	Read length (bp)	Reads	Aligned reads (Mac)	Aligned	Genome coverage %	Average Coverage
PGM	341	108	130 266 728	110 189 786	84,59%	99	161x
EZL #1	159	74	113 316 836	89 872 888	79,31%	99	89x
EZL #2	328	74	102 580 084	81 824 588	79,77%	99	83x
KLEB	212	74	79063468	77712438	98,29%	99	79x
DCL23	318	101	99744888	92247446	92,48%	99	130x

Table S3. Sequencing and mapping statistics. Read statistics are provided for the samples sequenced for this study, with the exception of the PGM silencing sample which was previously published (Arnaiz et al., 2012; European Nucleotide Archive Acc No. ERA137444).

Reference	PGM	EZL1	DCL23	KLEB
PGM	88996911 bp 100 %	88959057 bp 99,96 %	88430111 bp 99,36 %	76088297 bp 85,50 %
PGM not KLEB	12908614 bp 100 %	12870760 bp 99,71 %	12349370 bp 95,67 %	0 bp 0 %

Table S4. Sequence complexity of control, PGM, EZL1 and DCL2/3 datasets. The previously published contigs assembled from a PGM dataset (Arnaiz et al., 2012) were used as reference, representing the currently best available germline DNA assembly. However, only contigs larger than 1 kb were considered, representing 91 Mb of sequence complexity, of which 89 MB are covered by read mapping above our cutoff (i.e. 2 reads per kb of contig per million mapped reads in the library). Reads from each sample were mapped to the PGM contigs using BWA with default parameters, to determine the complexity of the contigs covered by at least 2 reads per kilobase of contig per million reads in the library (RPKM), giving the first row of the table ("PGM" Reference). In addition, a set of PGM contigs were selected that had control (KLEB) coverage below the cutoff of 2 RPKM ("PGM not KLEB" Reference), representing pure germline DNA not collinear with MAC chromosomes. The coverage of these contigs by each sample was also determined (second row of table). Note that 76 Mb is the complexity of the MAC reference genome (Aury et al., 2006) and that the total germline complexity is at least 10 Mb greater than 91 Mb. Although the N50 of the PGM assembly that we have used as Reference is 28,076 bp (meaning that half of the assembly is contained in contigs larger than 28 kb), the analysis presented in the table uses the 7,310 contigs greater in size than 1 kb out of a total of 30,013 contigs, a choice dictated by the necessity of obtaining good paired-end read mapping to calculate coverage.

NAME	Length (bp)	DCL2/3 RNAi			EZL1-RNAi		PGM-RNAi		Control	
		mcIES	PCR	Ret. score	PCR	Ret. score	PCR	Ret. score	PCR	Ret. score
IES51A1835	28	-	-	0	-	0.03636	+	0.80478	-	NA
IES51G1832	30	-	-	0.02128	-	0.05263	+	0.7337	-	0
IES51A4404	77	-	-	0.01316	-	0.05376 *	+	0.8287 ***	-	NA
IES51G-11	43	NA	-	0.0396	+	0.48421 ***	+	0.79512 ***	-	0
IES51G1413	52	-	-	0.00943	+	0.24468 **	+	0.78307 ***	-	0
mtA	195	+	+	NA	+	NA	+	NA	+	NA
IES51G2835	229	+	+	0.14706 *	+	0.42222 ***	+	0.77193 ***	-	0.00538
IES51A2591	370	+	+	0.43373	+	0.52482	+	0.9102	-	NA
IES51A6649	370	+	+	0.38017	+	0.45113 *	+	0.81333 *	-	0
IES51A4578	883	-	+/-	0.06329	+	0.45882	+	0.77844	-	NA

Table S5. IES retention analyzed by PCR and deep-sequencing after control, DCL2/3, EZL1 and PGM silencing. Maternally (mcIES) or non-maternally controlled (non-mcIES) is indicated as “+” or “-” respectively, based on previous studies (Duharcourt et al., 1998). Star(s) following a retention score indicate that the IES is significantly retained (cf. Materials and Methods and the legend to Figure 5B). NA: not analyzed.

ANCHOIS ID	SIZE	PGM	EZL1	EZL1_r	DCL23	DCL23_r	KLEB
IESPGM.PTET51.1.103.177611	2462	0.82379 - *	0.47959 - *	0.49682 - *	0.04132 - NS	0.07692 - NS	0
IESPGM.PTET51.1.104.49056	2238	0.82759 - *	0.53153 - *	0.54918 - *	0.32353 - *	0.48352 - *	0
IESPGM.PTET51.1.105.239361	2483	0.84783 - *	0.6 - *	0.5 - *	0.26126 - *	0.32927 - *	0
IESPGM.PTET51.1.120.182371	1769	0.83333 - *	0.5678 - *	0.42308 - *	0.15306 - *	0.11111 - NS	0.00685
IESPGM.PTET51.1.128.254421	3392	0.88199 - *	0.68182 - *	0.73171 - *	0.39604 - *	0.75728 - *	0
IESPGM.PTET51.1.132.167159	1500	0.75281 - *	0.55556 - *	0.5 - NS	0.1413 - NS	0.32812 - NS	0
IESPGM.PTET51.1.163.702	1956	0.82353 - *	0.59091 - *	0.5122 - *	0.19231 - *	0.47761 - *	0
IESPGM.PTET51.1.169.56908	3272	0.86447 - *	0.55556 - *	0.46667 - *	0.01887 - NS	0.04396 - NS	0
IESPGM.PTET51.1.173.70900	2714	0.93431 - *	0.95 - *	0.88235 - *	0.34783 - NS	0.67647 - *	0
IESPGM.PTET51.1.174.130670	3001	0.83908 - *	0.5098 - *	0.43011 - *	0.39735 - *	0.76647 - *	0
IESPGM.PTET51.1.181.1750	3001	0.57576 - *	0.4875 - *	0.39474 - *	0.05085 - NS	0.03636 - NS	0
IESPGM.PTET51.1.214.11549	2317	0.78161 - *	0.56962 - *	0.56452 - *	0.29496 - *	0.80303 - *	0
IESPGM.PTET51.1.24.100577	1340	0.81928 - *	0.51064 - *	0.40449 - *	0.2314 - *	0.30263 - *	0.01053
IESPGM.PTET51.1.28.457973	2468	0.84577 - *	0.48718 - *	0.6 - *	0.20213 - *	0.22951 - *	0
IESPGM.PTET51.1.29.290535	1722	0.74214 - *	0.54321 - *	0.45545 - *	0.33588 - *	0.78102 - *	0
IESPGM.PTET51.1.35.111752	2820	0.72131 - *	0.42424 - *	0.45455 - *	0.04511 - NS	0.0396 - NS	0
IESPGM.PTET51.1.42.397702	2473	0.84231 - *	0.56 - *	0.50667 - *	0.04167 - NS	0 - NS	0
IESPGM.PTET51.1.42.72890	3389	0.84135 - *	0.58209 - *	0.40541 - *	0.28986 - *	0.66667 - *	0
IESPGM.PTET51.1.47.408041	2125	0.83525 - *	0.61702 - *	0.53846 - *	0.39167 - *	0.67153 - *	0
IESPGM.PTET51.1.50.348282	1251	0.85714 - *	0.61667 - *	0.58333 - *	0.47761 - *	0.72436 - *	0
IESPGM.PTET51.1.51.131273	2219	0.85926 - *	0.54945 - *	0.47059 - *	0.23256 - *	0.34314 - *	0.01266
IESPGM.PTET51.1.57.48117	1257	0.78212 - *	0.36232 - *	0.64 - *	0.08333 - NS	0.16667 - NS	0
IESPGM.PTET51.1.76.220822	3048	0.83857 - *	0.65152 - *	0.5 - *	0.38255 - *	0.70526 - *	0
IESPGM.PTET51.1.77.209216	4154	0.87264 - *	0.56579 - *	0.63158 - *	0.33071 - *	0.5873 - *	0
IESPGM.PTET51.1.77.311405	3470	0.83682 - *	0.47059 - *	0.45833 - *	0.32117 - NS	0.60204 - *	0.05128
IESPGM.PTET51.1.80.84925	1513	0.79227 - *	0.53608 - *	0.52778 - *	0.12712 - *	0.25773 - *	0
IESPGM.PTET51.1.85.45587	3479	0.84965 - *	0.61702 - *	0.55844 - *	0.29167 - *	0.50485 - *	0
IESPGM.PTET51.1.98.309432	2003	0.88776 - *	0.62821 - *	0.72619 - *	0.61667 - *	0.80405 - *	0.01136

Table S6. Anchois retention score. The 28 IESs that were used to identify the Anchois Tc1/mariner transposon (Arnaiz et al., 2012) are provided with their ParameciumDB Accession Numbers, size and retention scores in the different RNAi datasets under consideration. Retention scores followed by a star are statistically significant (cf. Materials and Methods). The DCL23_r sample is a biological replicate previously published by Sandoval et al., 2014 that was retrieved from the Genbank Short Read Archive (Accession number SRX387766).

Table S7. Oligonucleotides used in this study.

Name	Sequence (5' to 3')	Application
bigNd7_5'	GATATACGTAAAATGATTGAATTGGTTTAC	PCR around ND7 gene
bigNd7_3'	TTCATAAGTTAACTTCTTAACTGATAATGC	PCR around ND7 gene
tel51G-01up	GATGTTAGACATATTAAGAGTTTGTTAGAC	tel51G probe
tel51G-02down	CTTAAGAATAAAAGCTATGCCAAAAGGAAGG	tel51G probe
EZL1_ATG_F	AAAAGCACCTAAATGACCAAAGAATAAAT	RT-PCR EZL1
EZL1_R5	CAATATCCTCTGTTAATTTGGTTCGG	RT-PCR EZL1
EZL2_U1_S	GGACTAGTATTACAGCATCCTGTCTATAAGATTTACTG	RT-PCR EZL2
EZL2_R1_S	GGACTAGTCTCTATAATGTAGCTTTCTTCATCCAACAAC	RT-PCR EZL2
EZL3a_U1	GCTCTTGTGAATTATCCATTTAATAATGC	RT-PCR EZL3a
EZL3a_R1	ATGAAAATATTTGCCTGGTTTTGCAATTC	RT-PCR EZL3a
EZL3b_U2	TAAGTAGGTATCGATTAAATGATTCTG	RT-PCR EZL3b
EZL3b_R2	TTAGAATATTGTTTGCTTTATATGAGAA	RT-PCR EZL3b
EZL4_U1_S	GGACTAGTCGCATTAATTAATAATCCATATAAAATCTG	RT-PCR EZL4
EZL4_R1_S	GGACTAGTCAAGTAAAAATTTGATTTGTTCTACAATTA	RT-PCR EZL4
PGM_F2	GAAAGAGGTACAGGAATGTGAATTC	RT-PCR PGM
PGM_R2	GCAGTTTTCGGATCATCAAACAAGGATC	RT-PCR PGM
T1b_5' (2)b	TCTAATTAAACCAAGAACACGCTGAATTC	RT-PCR T1b
T1b_3'	TTGAGTTGGGATTTGACATAATCGGTGAA	RT-PCR T1b
51A4578-2	TGGTTGTTAGTCTCAAAGAATTCTAAAGAC	PCR for IES circles
51A4578-7	AAGAAATTTTATTGTAAATATATTTTCAGC	PCR for IES circles
51A-1835-5' (3)	GGTTGCGTAACACTTCCTCTTAAATGTGAG	PCR around IESA1835
51A-1835-3'	ATCCTAACATCCTTGAATAGTTACTGATCC	PCR around IESA1835
51G1832up	GCTATAACTCTTGAAGCTGCTTGTAATATG	PCR around IESG1832
51G1832do	TTGTCAATGAGCCATTAACAGTTGCTGGAT	PCR around IESG1832
51A-4404-5' (2)	TGGAATAGTGCTGCATCACCAGCTGCTTGC	PCR around IESA4404
51A-4404-3' (2)	CCAGTTATTGAACTGCAACTTACTGCAGTG	PCR around IESA4404
51G-11up	ATCATAAGATTGATATCTTCTCCCTTCTCC	PCR around IESG-11
51G-11do	ACTTGCTACTAAAGCAAGAAACATTGAGAG	PCR around IESG-11
51G1413up	GAAGCTGCTTGTGTTAAGAATTCTACTGG	PCR around IESG1413
51G1413do	GCATCCAGCACTAGTTGAATTTACTGTA	PCR around IESG1413
mtAF6	GGTGTTTTATATCTTAATTGTTGACCCTCAC	PCR around IES mtA
mtAR7	CCATCTATACTCCATTCTTTATCTTAATTCAT	PCR around IES mtA
51G2832-6	CAAATACTGGTGGTGCAACATCCTCAACTGC	PCR around IESG2835
51G2832-8	GAACCATTTAATGCGCAATATCCTGCATTTCC	PCR around IESG2835
51A2591-18	AAGTGCAACCTGTGCTGATGCTCCCGATGA	PCR around IESA2591
51A2591-20	AGTTCCTTTGAAAGATGTGCAAGCTCCAGA	PCR around IESA2591
51A-6649-5'	AAATGGTACTGTTTGTGCTTGGGATAGTGC	PCR around IESA6649
51A-6649-3'	CAGCAGTACATCCAGCTCTCTAAGTTTAGC	PCR around IESA6649
51A4578-5'	CACTGCAGTAAGTTGCAGTTCAATAACTGG	PCR around IESA4578
51A4578-3'	GTAGTCTTAAATCTTAGCATGTTGTACC	PCR around IESA4578

RESULTATS COMPLEMENTAIRES

Analyse à l'échelle du génome entier de la rétention des IES dans des cellules inactivées pour PTIWI01 et PTIWI09

Nous avons montré que l'inactivation de EZL1 conduisait à la rétention d'un sous-ensemble d'IES à la fin du développement macronucléaire. De plus, l'inactivation des gènes DCL2 et DCL3 conduit à la rétention d'un sous-ensemble d'IES sensibles à EZL1. Certaines IES sont donc sensibles à EZL1 mais insensibles à DCL2 et DCL3 suggérant qu'il existe une voie d'excision dépendante de la méthylation des histones mais indépendante des scanARN. Il pourrait exister d'autres petits ARN responsables du ciblage de Ezl1p au niveau des séquences à éliminer. Notamment, les protéines Piwi, qui possèdent une activité « slicer », peuvent produire des petits ARN indépendamment des protéines Dicer (Parker and Barford, 2006). Chez *Paramecium*, deux protéines Piwi, Ptiwi01 et Ptiwi09 interviennent spécifiquement pendant le développement macronucléaire et sont requises pour l'accumulation des scanARN produits par Dcl2 et Dcl3 (Bouhouche et al., 2011). D'après Bouhouche *et al.*, la déplétion des protéines Ptiwi01 et 09 conduit à la rétention de toutes les IES testées (une dizaine) dans le macronoyaux en développement. L'élimination des IES insensibles à DCL2 et DCL3 pourrait donc être médiée par des petits ARN produits par les protéines Ptiwi.

Pour valider cette hypothèse à l'échelle du génome entier, nous avons séquencé le génome macronucléaire de cellules inactivées pour PTIWI01 et PTIWI 09 avec la même procédure que pour les génomes des cellules inactivées pour EZL1 ou DCL2/3. Les plasmides utilisés pour l'inactivation sont les mêmes que ceux décrits dans Bouhouche *et al.* Dans notre expérience, seulement 60% de létalité (20 cellules/30) a été observée dans la progéniture sexuelle des cellules inactivées pour PTIWI01 et PTIWI09. L'inactivation d'autres gènes impliqués dans les réarrangements du génome conduit généralement à 90-100% de létalité (Baudry et al., 2009; Lepère et al., 2008; Nowacki et al., 2005; Nowak et al., 2011). Cependant, la létalité que nous avons observée dans cette expérience de co-inactivation de PTIWI01 PTIWI09 est cohérente avec

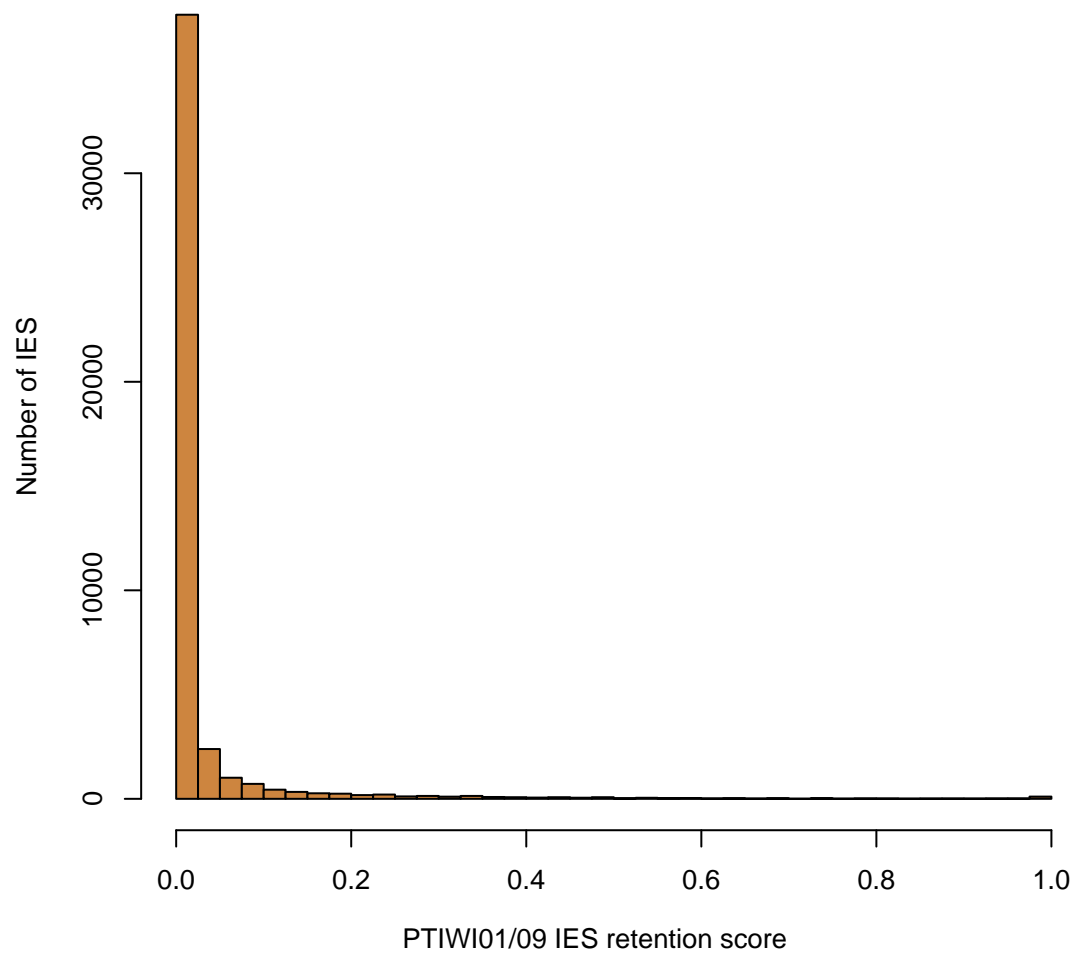


Figure 18. Distribution des scores de rétention des IES dans le génome des cellules inactivées pour PTIWI01 et PTIWI09 (Voir Lhuillier-Akakpo et al., pour la procédure de calcul des scores de rétention).

la létalité décrite dans les descriptions de précédentes (Bouhouche et al., 2011), où 8/12 cellules testées meurent (environ 60%).

L'analyse de la rétention des IES dans le génome de cellules inactivées pour PTIWI01/09 a été réalisée comme notre étude sur EZL1 (Lhuillier-Akakpo et al., *en préparation*). La distribution des scores de rétention des IES est présentée sur la Figure 18. Dans cette expérience, 1194 IES sont significativement retenues dans le génome des cellules inactivées pour PTIWI01/09. Le score moyen de rétention des IES est de 0,028. Parmi les 1194 IES sensibles, 1120 sont également sensibles à DCL2/3 et 74 y sont insensibles. Il reste donc 2152 IES sensibles à DCL2/3 mais insensibles à PTIWI01/09.

La rétention des IES après l'inactivation de PTIWI01/09 est donc très faible. Le fait que toutes les IES que nous avons identifiées comme sensibles à DCL2/3 ne sont pas sensibles à PTIWI01/09 suggère que l'inactivation par ARN interférence des gènes PTIWI01 et 09 n'a pas été complète. En effet, les protéines Ptiwi sont requises pour l'accumulation des scanARN et donc les IES sensibles à PTIWI01/09 devraient au moins correspondre au même groupe que les IES sensibles à DCL2/3.

De plus, nos résultats semblent en contradiction avec les données publiées qui décrivaient, sur un petit échantillon, que les protéines Ptiwi étaient requises pour l'élimination de toutes les IES (Bouhouche et al., 2011). Il faut cependant noter que dans l'étude qui avait été réalisée, la rétention des IES dans le macronoyau en développement est analysée assez précocement (autour de 15 heures après le début des événements sexuels). Dans nos expériences, la rétention des IES est analysée bien plus tard, autour de 70-80 heures après le début des événements sexuels, quand le développement est achevé. Dans Bouhouche et al., les deux seules IES (G4404 et G2832) qui sont montrées comme retenues à des stades aussi tardifs le sont aussi dans notre expérience (non montré). La rétention des IES observée à 15 heures dans les cellules inactivées pour PTIWI01 et 09 pourrait donc correspondre à un retard d'élimination et non à une rétention définitive, expliquant les différences avec nos données.

Il est donc difficile d'interpréter les résultats de notre expérience. Le faible phénotype observé pourrait provenir d'une combinaison entre un effet faible de l'inactivation et un effet faible des protéines Ptiwi dans le processus d'élimination d'ADN. Il nous faudra reproduire cette expérience

d'inactivation de PTIWI01 et 09 plusieurs fois et séquencer le génome des cellules inactivées pour conclure sur le rôle de ces gènes dans l'élimination des IES.

PARTIE 2

Mécanismes d'inactivation programmée des centromères au cours du développement macronucléaire de *Paramecium tetraurelia*

ARTICLE 2 (en préparation)

Chez le cilié *Paramecium tetraurelia*, micronoyaux et macronoyau se distinguent à la fois par leur contenu génétique et par leur activité centromérique. Tandis que les micronoyaux ségrègent par mitose classique pendant la croissance végétative des cellules, le macronoyau se divise par un mécanisme non mitotique qui n'impliquerait ni condensation des chromosomes, ni formation de fuseau mitotique (Tucker et al., 1980). Les chromosomes macronucléaires se comportent donc comme s'ils étaient dépourvus de fonction centromérique.

Dans l'étude qui va suivre, nous avons cherché et identifié le variant d'histone centromérique CenH3 de *P. tetraurelia* et montré que la fonction centromérique est restreinte aux chromosomes micronucléaires. En utilisant CenH3 comme un marqueur des centromères actifs, nous avons montré que l'inactivation des centromères germinaux se produit au cours du développement du macronoyau, au moment même où a lieu l'élimination massive d'ADN micronucléaire. Pour comprendre les mécanismes qui sont impliqués dans l'inactivation des centromères chez *Paramecium*, nous avons testé le rôle de gènes impliqués dans les réarrangements du génome pendant le développement macronucléaire. Nous avons pu observer que l'inactivation des centromères dépendait de l'endonucléase PiggyMac et de la lysine méthyltransférase Ezl1p suggérant fortement que l'inactivation des centromères se produit par l'élimination des séquences centromériques.

Programmed centromere inactivation in the ciliate *Paramecium tetraurelia*

Maoussi Lhuillier-Akakpo^{1,2}, Frédéric Guérin¹, Jérôme Poli & Sandra Duharcourt¹

¹Institut Jacques Monod, CNRS, UMR 7592, Université Paris Diderot, Sorbonne Paris Cité, Paris, F-75005 France.

²Sorbonne Universités, UPMC Univ., IFD, 4 place Jussieu, 75252 Paris cedex05, France.

Correspondence should be addressed to S. D. (duharcourt@ijm.univ-paris-diderot.fr)

RESULTS

Two centromeric histone variants are present in *Paramecium* genome

To identify centromeric histones, we looked for H3 proteins and H3 variants in the *Paramecium* macronuclear genome by sequence homology with the human canonical H3 protein (Figure 1). We found three canonical histones H3 named H3P1, H3P12 and H3P3, encoded respectively by five genes, one gene and three genes (see Table S1 for accession numbers). Two other proteins, encoded by single genes, presented features of H3.3 variants and were named H3P2 and H3P4. They are very similar to the canonical H3s but present a substitution in position 94 transforming a valine into an isoleucine, characteristic of H3.3 variants (Hake and Allis, 2006). We then identified three more divergent H3 variants, H3P5, H3M1 and H3M2. We have not been able to attribute a putative function to these proteins on the basis of their amino acid sequences. Since they have not been characterized yet, they will not be presented in this study. Finally, we found two H3 variants, CenH3-1 and CenH3-2, with all CenH3 sequence signatures: 1. a low conservation in the histone fold domain (43% of amino acid identity with H3P1), 2. a highly divergent N-terminal tail, 3. the presence of additional amino acids in the first loop of the histone fold domain. CenH3-1 and CenH3-2 are duplicates from the last whole genome duplication (Aury et al., 2006) and present 90,3% of amino acid identity. We examined CenH3-1 and CenH3-2 expression profiles during *Paramecium* life cycle, using RNA-sequencing data obtained by mRNA sequencing of cells at different stages of the life cycle (Arnaiz et al., in preparation). CenH3-1 is weakly expressed in vegetative cells but slightly up-regulated at the beginning of the sexual events. It is then down-regulated during the macronuclear development. CenH3-2 is less expressed than CenH3.1 in vegetative cells but is highly up-regulated at the beginning of the sexual events. It comes back to a basal level at the end of the macronuclear development (Figure 2).

CenH3-1 is required for the micronuclear functions

To test whether the putative CenH3 genes of *Paramecium* are functional and can play the role of centromeric histone variants, we observed the phenotypes associated to CenH3-1 and CenH3-2 depletion during the vegetative growth. In eukaryotic cells, depletion of kinetochore proteins leads to chromosome segregation defects in mitosis, cell cycle arrest and cell death (Blower and Karpen, 2001; Stoler et al.,

1995). In *Paramecium* cells, CenH3 silencing should impair the functions of the nuclei which require active centromeres. Alteration of macronuclear functions will lead to cellular growth defects whereas alteration of micronuclear functions will impair the viability of the sexual progeny. CenH3-1 and CenH3-2 were silenced independently by RNAi. Cells were fed with bacteria engineered to produce double stranded RNA carrying a specific portion of CenH3-1 or CenH3-2 genes. Double stranded RNAs will be processed by the cellular machinery into siRNAs that will degrade the messenger RNAs of the targeted genes. To silence CenH3-1 and CenH3-2 specifically and independently, we chose gene fragments in the N-terminal region, the most divergent portion between the two sequences (see material and methods for gene fragment positions). No growth defect was observed after 3 days of vegetative divisions compared to control cells, indicating that CenH3-1 and CenH3-2 genes are not required for the macronuclear functions (not shown). To assess whether CenH3 genes silencing affected the micronuclei, cells were fed back in a normal medium to stop the silencing process and starved to trigger the sexual cycle. Strikingly, cells that had been silenced for CenH3-1 during the vegetative cycle were not able to produce a viable sexual progeny, even after only one day of vegetative divisions (Table 1). Therefore, CenH3-1 is required for micronuclear functions during the vegetative cycle. Conversely, the silencing of CenH3-2 during the vegetative cycle had no lethal effect on the sexual progeny suggesting that CenH3-2 has no essential micronuclear function during the vegetative growth (Table 1). However, CenH3-2 expression profile sustains the possibility of a role during the sexual cycle or the macronuclear development. To verify this hypothesis, we silenced each CenH3 gene at the onset of autogamy. Cells were grown for 4-5 divisions and starved in the silencing medium to trigger the sexual cycle in RNAi conditions. CenH3-1 silencing during the sexual cycle led to a strong lethality of the sexual progeny (Table 2). Since cells undergo few vegetative divisions in the silencing medium before starting the sexual events, the effects of CenH3-1 silencing during the sexual cycle, even if they might exist, cannot be distinguished from the vegetative effects (observed in Table 1). Although CenH3-2 mRNAs are upregulated at early stages of autogamy, CenH3-2 silencing during autogamy did not impair the viability of the sexual progeny (Table 2) suggesting that this gene is not essential during the sexual cycle.

Macronuclear chromosomes are deprived of active centromeres

Because CenH3-1 is required for micronuclei functions, it suggests that only micronuclear chromosomes carry a centromeric function. To confirm this hypothesis, we monitored the subcellular localization of the endogenous CenH3-1 protein. As CenH3 proteins are highly divergent between species, we produced a polyclonal antibody specific for CenH3-1 of *Paramecium tetraurelia* from a peptide chosen in the N-terminal region of CenH3-1 between positions 5 and 24 (KTTKENNNQSFQVDNNEKMP). We performed immunostaining experiments during *Paramecium* vegetative divisions. As expected, CenH3-1 localizes only in micronuclei and is absent from the macronucleus at all stages of the cell cycle (Figure 3A). This localization is consistent with the one observed with a CenH3-GFP fusion protein (Figure 3B) confirming the specificity of our antibody. Unfortunately, micronuclear chromosomes are very small, and this makes extremely difficult to observe CenH3 localization at the chromosomal level. Using CenH3 as a marker of active centromeres, we concluded that, in *Paramecium* cells, the centromeric function is restricted to the germinal nuclei and that macronuclear chromosomes are deprived of active centromere.

Germline centromeres are inactivated during macronuclear development, concomitantly with DNA elimination events

Since micronuclei and macronuclei differentiate from mitotic products of the zygotic nucleus, we performed immunostaining experiments with the CenH3 antibody during the early sexual events on conjugating cells to determine at which developmental stage the centromeric function is lost from the macronucleus. We observed that CenH3 localized in the micronuclei during meiosis I and meiosis II and was present in the eight haploid gametic nuclei (Figure 4A, panels a-c). CenH3 was still present in the zygotic nucleus after karyogamy but at this stage had disappeared from the seven other haploid nuclei that will eventually degenerate (Figure 4A, panel d). The centromeric protein was found in the four diploid products resulting from the two mitotic divisions of the zygotic divisions. These four diploid nuclei will then differentiate into the new micronuclei and the new macronuclei (Figure 4A panel e, Figure 4B panel a). We then followed CenH3 staining during the macronuclear development of autogamous cells, whose progression can be monitored by the progressive swelling of the two new macronuclei, due to DNA endoreplication.

CenH3 was found in the new micronuclei and in the new macronuclei at early stages of development (Figures 4B, panels b-c) but disappeared from the two developing macronuclei at later stages (Figure 4B, panels d-e). Since CenH3 is a marker of active centromere, CenH3 disappearance in the new macronuclei reflects a process of centromere inactivation and explains the absence of centromeric function in the vegetative macronucleus.

To determine precisely the moment at which centromere inactivation occurs relatively to DNA rearrangement events, we performed a developmental time course experiment in which macronuclear CenH3 staining has been followed concomitantly with DNA elimination events. The proper progression of autogamy in the cell population was monitored by nuclei coloration of cell samples taken at different time points. Cells do not enter in the sexual cycle synchronously, then, each time point corresponds to an heterogeneous population of cells that are at different stages of the sexual cycle. At time 0 hr, about 80% of cells have started autogamy and we could observe a rapid apparition of cells with two new macronuclei (50% after 5 hrs and 92% after 15 hrs), due normal progression of the autogamy process in the cell population (Figure 5A). Among cells containing two developing macronuclei, the proportion of CenH3 positive macronuclei was counted. Most cells are CenH3 positive in the first hours of the macronuclear development (82% to 95% between 5 hrs and 20 hrs). Yet, this proportion strongly fell down to 0% between 20 hrs and 35 hrs after the beginning of the sexual events (Figure 5B). This severe drop might reflect that centromere inactivation relies on an active process. The moment of centromere inactivation is concomitant with the apparition of karyonides, cells that have realized the first cellular division after the sexual cycle. However, CenH3 signal has been lost in both karyonides and cells containing two new macronuclei, showing that centromere inactivation is a process independent from the cellular division. To relate centromere inactivation to DNA elimination, we followed in the same developmental time course experiment the expression profile of PiggyMac, the putative endonuclease required for the introduction of double strand breaks at the extremities of germline-limited sequences (Baudry et al., 2009). PiggyMac expression was induced at 0 hr and reached a maximal level between 10 hrs and 20 hrs, a moment that precedes centromere inactivation (Figure 5C). We detected excision products for IES A4578 known to circularize after its excision from

macronuclear chromosomes (Bétermier et al., 2000). Our PCR results revealed that A4578 IES excision products can be detected between 10 hrs and 20 hrs and were no longer detected at T35 hrs, consistent with the existence of an active mechanism involved in the degradation of these DNA elements between 20 hrs and 35 hr (Figure 5C). Finally, imprecise DNA elimination was monitored by PCR for the ND7 gene, which, in this cell line, is deleted from the MAC and thus only present in the MIC (described in (Garnier et al., 2004)). In the first stages of the sexual cycle (0 hr – 5 hrs), the macronuclear deleted form of the ND7 gene was revealed as a smear because of the presence of multiple deletion boundaries. During the progression of the macronuclear development, the micronuclear and complete form of the ND7 gene was detected due to its amplification in the new macronuclei, before being deleted between 20 hrs and 35 hrs, concomitantly with CenH3 and IES circles disappearance (Figure 5C). Together, these results suggest that centromere inactivation is an active process that occurs in the late stages of the macronuclear development, in the same developmental time window as degradation of IES excision products and imprecise DNA elimination events.

Centromere inactivation depends on genes involved in the DNA rearrangements

The timing of centromere inactivation in the developing macronuclei is consistent with the hypothesis that centromeric sequences are eliminated from germline chromosomes during the process of genome rearrangements. Therefore we tested whether the putative endonuclease PiggyMac is involved in centromere inactivation. RNAi-mediated silencing of PiggyMac was performed during the sexual cycle in three independent experiments and led, as expected, to a strong lethality of the sexual progeny (Baudry et al., 2009). The proportion of cells containing CenH3 positive macronuclei was followed in control and in PiggyMac silencing conditions by immunofluorescences with the CenH3 antibody. Two different stages of the macronuclear development are shown on Figure 6: an early stage, between 15 and 20 hrs after the beginning of the sexual cycle, when CenH3 is present in the macronuclei of control cells and a later stage, between 40 and 50 hours, when CenH3 has normally disappeared from the new macronuclei. PiggyMac silencing did not alter the presence of CenH3 in the early developing macronuclei but impaired its disappearance at the end of macronuclear development. Indeed, 100% of observed

cells in each experiment retained the CenH3 signal at a late stage of macronuclear development whereas CenH3 had completely disappeared in control cells. This result strongly supports the hypothesis that centromere inactivation is due to the physical elimination of centromeric sequences. We next silenced the putative histone methyltransferase EZL1, responsible for the developmentally specific trimethylation of H3K27 and H3K9 in the developing macronuclei and that is required for the imprecise elimination of mic-specific sequences and the precise excision of a subset of IESs (Lhuillier-Akakpo et al., *in preparation*). As observed for PiggyMac, EZL1 silencing prevented centromere inactivation, as CenH3 signal was maintained at late stages of macronuclear development in three independent experiments (Figure 6). Therefore, centromere inactivation might depend on H3K9 and H3K27 methylation. If centromeres are mic-specific sequences, then centromeric sequences should belong to the subset of EZL1-sensitive eliminated sequences. Finally, we silenced DCL2 and DCL3, the two genes involved in the biogenesis of scanRNAs (Lepère et al., 2009). DCL2 and DCL3 depletion impairs imprecise elimination of mic-specific sequences and the precise excision of a subset of EZL1-sensitive IESs (Lepère et al., 2009) ; (Lhuillier-Akakpo, *in preparation*). We observed that DCL2 and DCL3 silencing did not affect CenH3 disappearance from the new macronuclei at the end of the development, showing that centromere inactivation does not depend on scanRNAs (Figure 6). This result suggests that centromeric DNA sequences are eliminated by a pathway dependent on PGM and EZL1 but independent of the scanRNAs.

Table 1. Viability of the sexual progeny after RNAi silencing of CenH3-1 and CenH3-2 during few vegetative divisions

	Silencing in vegetative growth			
	no silencing	control	CenH3-1	CenH3-2
viable (%)	100	100	0	100
total nb of cells	36	132	110	36
nb of experiments	2	2	2	2

no silencing: cells have been grown in a normal Klebsiella medium

control: inactivation of a non relevant gene

Table 2. Viability of the sexual progeny after RNAi silencing of CenH3-1 and CenH3-2 during autogamy

	Silencing during autogamy			
	no silencing	control	CenH3-1	CenH3-2
viable (%)	94	99	19	94
total nb of cells	120	144	48	96
nb of experiments	3	3	1	2

no silencing: cells have been grown in a normal Klebsiella medium

control: inactivation of a non relevant gene

		9		27	
H3P1	1	-MARTKQTAR	KSTA	-----GNKKPTKHLATKAARK	T
H3P12	1	-MARTKQTAR	KSTT	-----GNKKPTKHLATKAARK	T
H3P2	1	-MARTKQTAR	KSTA	-----GNKKPTKHLATKAARK	T
H3P3	1	-MARTKQTAR	KSTA	-----GNKKPTKHLATKAARK	T
H3P4	1	-MARTKQTAR	KSTA	-----GNKKPTKHLATKAARK	T
H3P5	1	-MARTKQ	GKPSNNK	-----FSLSPQ-SMCNDSL	--E
H3M1	1	-MARTKEG	-----	-----KEMSKKSI	----AKNS
H3M2	1	-MARTKEG	-----	-----KEMSKKSI	----AKTS
CenH3-1	1	-MAN	KKTTKENNNQSFQVDNNEKMPSFMHSLFDSDEKSVIAEKS	NE	SKKSEKKKVERIS
CenH3-2	1	MAN	KKKAKENENQSFQVDNVEQMPSFMHSLFNSDEKSI	IAEKSKEESKKSERK	RDERIS
HsH3	1	-MARTKQTAR	KS	-T-----CAKAPRKQLASKAARK	S
H3P1	31	APAVGATGGLKKPHKFRPGTVALREIRKYQKSTELLIRKLPFQRLVREIAHEF	-----	QK	
H3P12	31	APAVGATGGLKKPHKFRPGTVALREIRKYQKSTELLIRKLPFQRLVREIAHEF	-----	QK	
H3P2	31	APAVGPTGGLKKPHKFRPGTVALREIRKYQKSTELLIRKLPFQRLVREIAHEF	-----	QK	
H3P3	31	APAVGAAGGLKKPHKFRPGTVALREIRKYQKSTELLIRKLPFQRLVREIAHEF	-----	QK	
H3P4	31	APAVGPSGGLKKPHKFRPGTVALREIRKYQKSTELLIRKLPFQRLVRDIAHDF	-----	QK	
H3P5	28	NKKQIYKNQOKRKHKFRPGTVSLREIRKYQSSTOLLIRKLPFQRLVREVSQRS	-----	EK	
H3M1	20	GKKMKSKRKERKPHRFRPGTVALKEIRKFOESSRLLIRKLPFQRLIREIAGVN	-----	EK	
H3M2	20	GKKMKSKRKERKPHRFRPGTVALKEIRKFOESSRLLIRKLPFQRLIREIAGVN	-----	EK	
CenH3-1	60	-AIQVQKARDKLQKRNPMSKVLQEI	RQLQASSVLVCRFAGFORFVROTGIKV	SDEL	GFK
CenH3-2	61	-GIQVQKARDKLQKRNPMSKVLQEI	RQLQTSGLVCRFAGFORFVROAGIKV	SDEL	GFK
HsH3	30	APA---	TGGIKKPHRFRPGTVALREIRKYQKSTDLLIRKLPFQRLVRDIAHEF	-----	KA
H3P1	86	ELRFQSSAVLALQEAAEAYLVGLFEDTNLCAIHARRVTIMSRDIQLARRIRGERF	-----		
H3P12	86	ELRFQSSAVLALQEAAEAYLVGLFEDTNLCAIHARRVTIMSRDIQLARRIRGERF	-----		
H3P2	86	ELRFQSSAILALQEAAEAYLVGLFEDTNLCAIHARRVTIMSRDIQLARRIRGERF	-----		
H3P3	86	ELRFQSSAVLALQEAAEAYLVGLFEDTNLCAIHARRVTIMSRDIQLARRIRGERF	-----		
H3P4	86	ELRFQSSAILALQEAAEGYLVGLFEDTNLCAIHARRVTIMSRDIQLARRIRGERF	-----		
H3P5	83	DFRFTQSALYALQEATEGFITHLFEDANLCSHAKRVTLLTRDLQLATRIQGI	EY		
H3M1	75	EMRFQSSAIFALQEATEAFIVNLLSDSVLCAHAKRMTVMORDINLARRIRGDEF	-----		
H3M2	75	EMRFQSSAIFALQEATEAFIVNLLSDSVLCAHAKRMTVMORDINLARRIRGDEF	-----		
CenH3-1	119	EFYYSKSL	ECLQTLTEQYMVDLFEDSVQCTFHAKRVTLM	AKDLNLTARIRGIEQPLQE	I
CenH3-2	120	EFYYSKSL	ECLQTLTEQYMVDLFEDSVQCTFHAKRVTLM	AKDLNLTARIRGIEQPLQE	I
HsH3	82	ELRFQSSAVLALQEAAEAYLVGLFEDTNLCAIHARRVTIMTKDMQLARRIRGERF	-----		
H3P1		-----			
H3P12		-----			
H3P2		-----			
H3P3		-----			
H3P4		-----			
H3P5		-----			
H3M1		-----			
H3M2		-----			
CenH3-1	179	RNLKLR			
CenH3-2	180	RNLKLN			
HsH3		-----			

Figure 1. Multiple sequence alignments of *Paramecium* H3 proteins and H3 variants

H3P1, H3P12 and H3P3 are canonical H3 proteins of *Paramecium tetraurelia*, homologs of the human histone H3 (HsH3). H3P2 and H3P4 are H3.3 variants. Lysines K9 and K27 of the N-terminal tail are indicated in red. Five more divergent histone variants have been found: H3P5, H3M1 and H3M2 have still unknown functions whereas CenH3-1 and CenH3-2 are two putative centromeric histone variants. CenH3-1 and CenH3-2 have very divergent N-terminal regions compared to canonical H3 proteins and contain additional amino acids in the first loop of the histone fold domain (indicated in red).

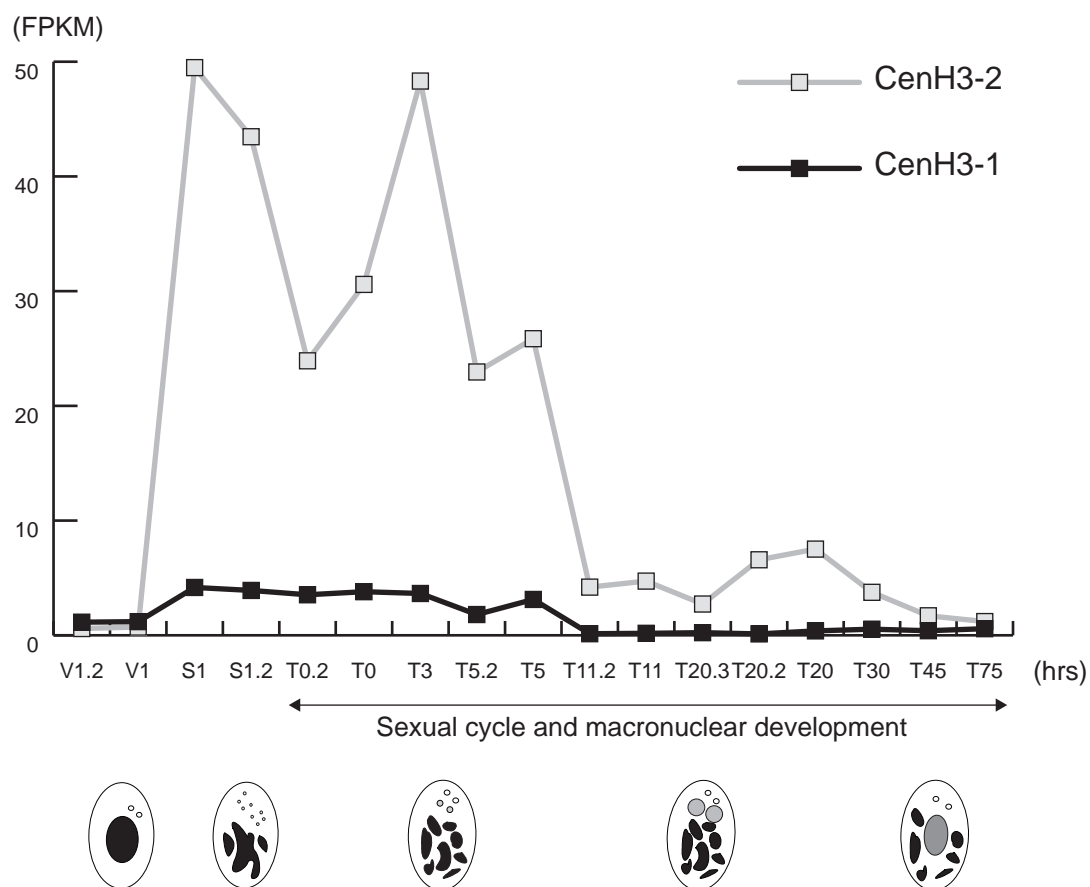


Figure 2. CenH3-1 and CenH3-2 expression profiles during *Paramecium* life cycle

RNA sequencing data (Arnaiz et al., in preparation) showing CenH3-1 (black line) and cenH3-2 (gray line) levels of expression in vegetative cells (V1.2 and V1), in starved cells (S1 and S1.2) and during the sexual cycle and the macronuclear development (T0 to T75 hours). Some time points are duplicated because RNAs coming from two independent developmental time courses have been sequenced. FPKM: fragments per kilobase per million

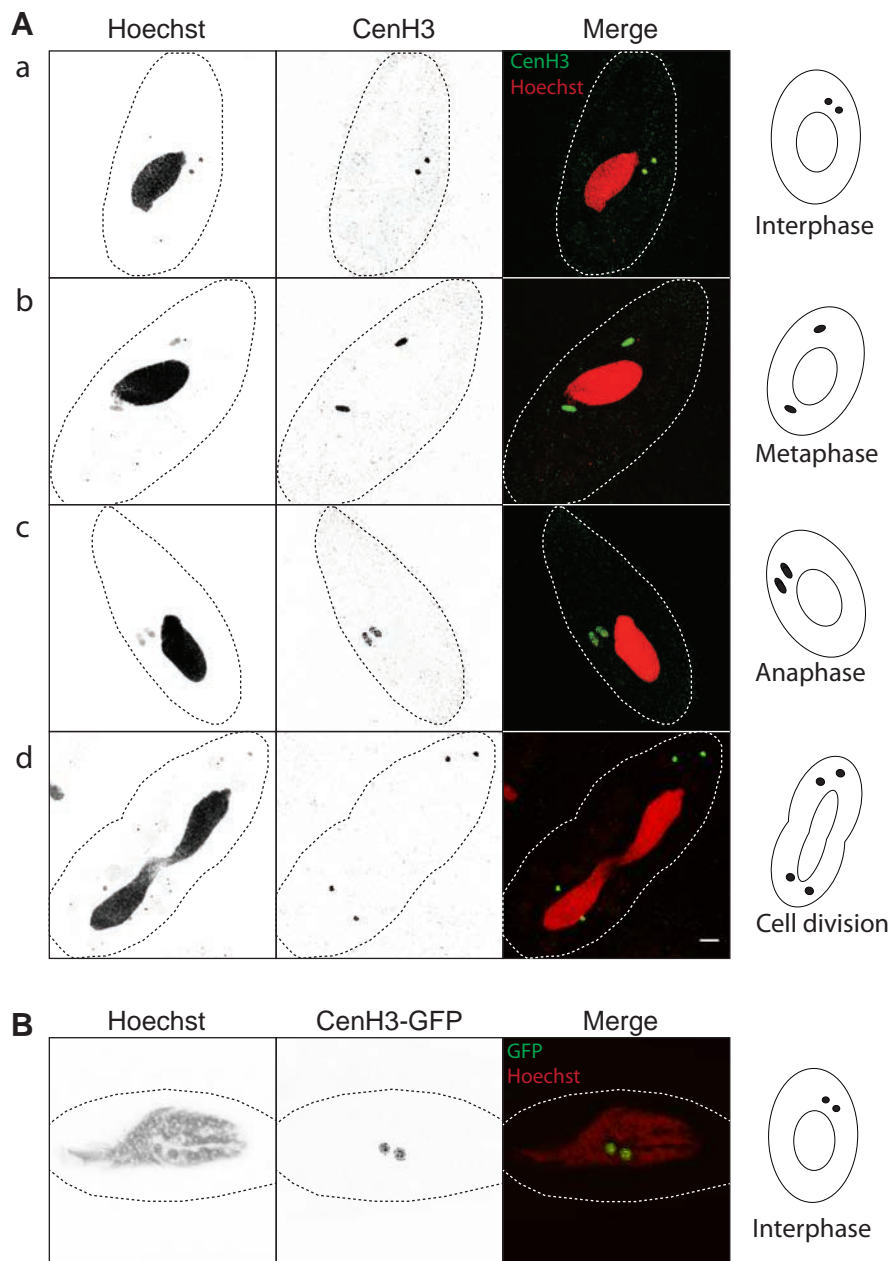


Figure 3. CenH3-1 localizes in micronuclei during the entire cell cycle of *Paramecium*

A. Immunostaining with a polyclonal anti-CenH3 antibody on *Paramecium* cells (a) in interphase, (b) metaphase, (c) anaphase and (d) during cell division. Pictures are z projections of several confocal sections. Dashed lines indicate the shape of cells. In the merge panel, red color corresponds to the hoechst fluorescence and green color corresponds to the CenH3 fluorescence. The scale bar indicates 10 μ m. **B.** Confocal section of an interphasic *Paramecium* cell expressing a CenH3-GFP fusion protein. CenH3-GFP localizes only in micronuclei.

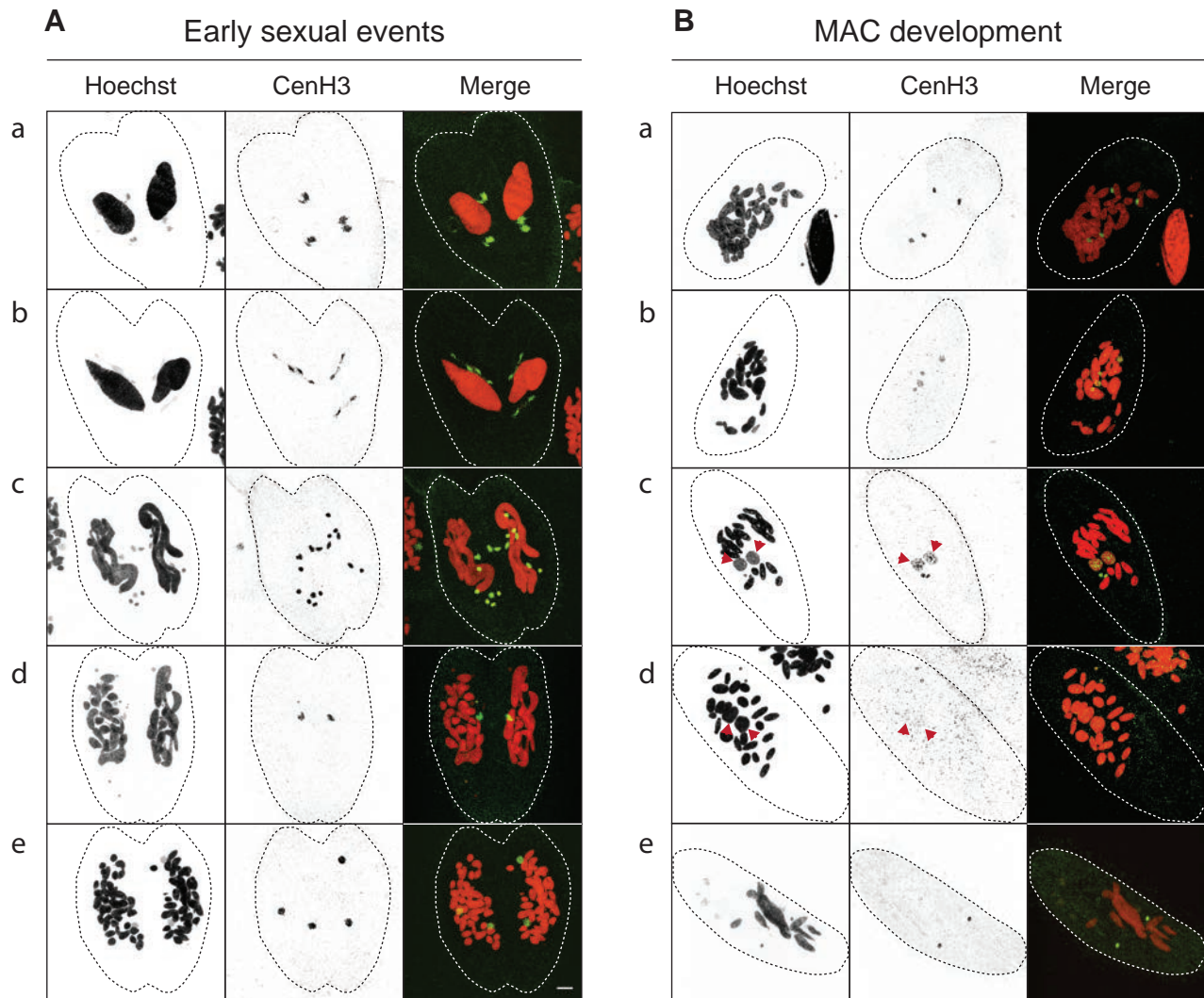


Figure 4. Centromere inactivation occurs during macronuclear development

A. Immunostaining with anti-CenH3 antibody on *Paramecium* conjugating cells allowing the observation of CenH3 localization during early sexual events. (a) Prophase of meiosis I. (b) Anaphase of meiosis I. (c) End of meiosis II. Eight haploid nuclei are present in cells. (d) karyogamy, CenH3 is present in the zygotic nucleus but has disappeared from the seven degenerating haploid nuclei. (e) First division of the zygotic nucleus. **B.** Immunostaining with anti-CenH3 antibody on autogamous *Paramecium* cells. (a) Second division of the zygotic nucleus. (b-c) Early macronuclear development. CenH3 is present in the new macronuclei and in the new micronuclei. (d) Late macronuclear development. CenH3 has disappeared from the developing macronuclei. (e) Karyonide. (c-d) Red arrows indicate the developing macronuclei. (a-e) Dashed lines represent the shape of the cells. The scale bar indicates 10 μ m.

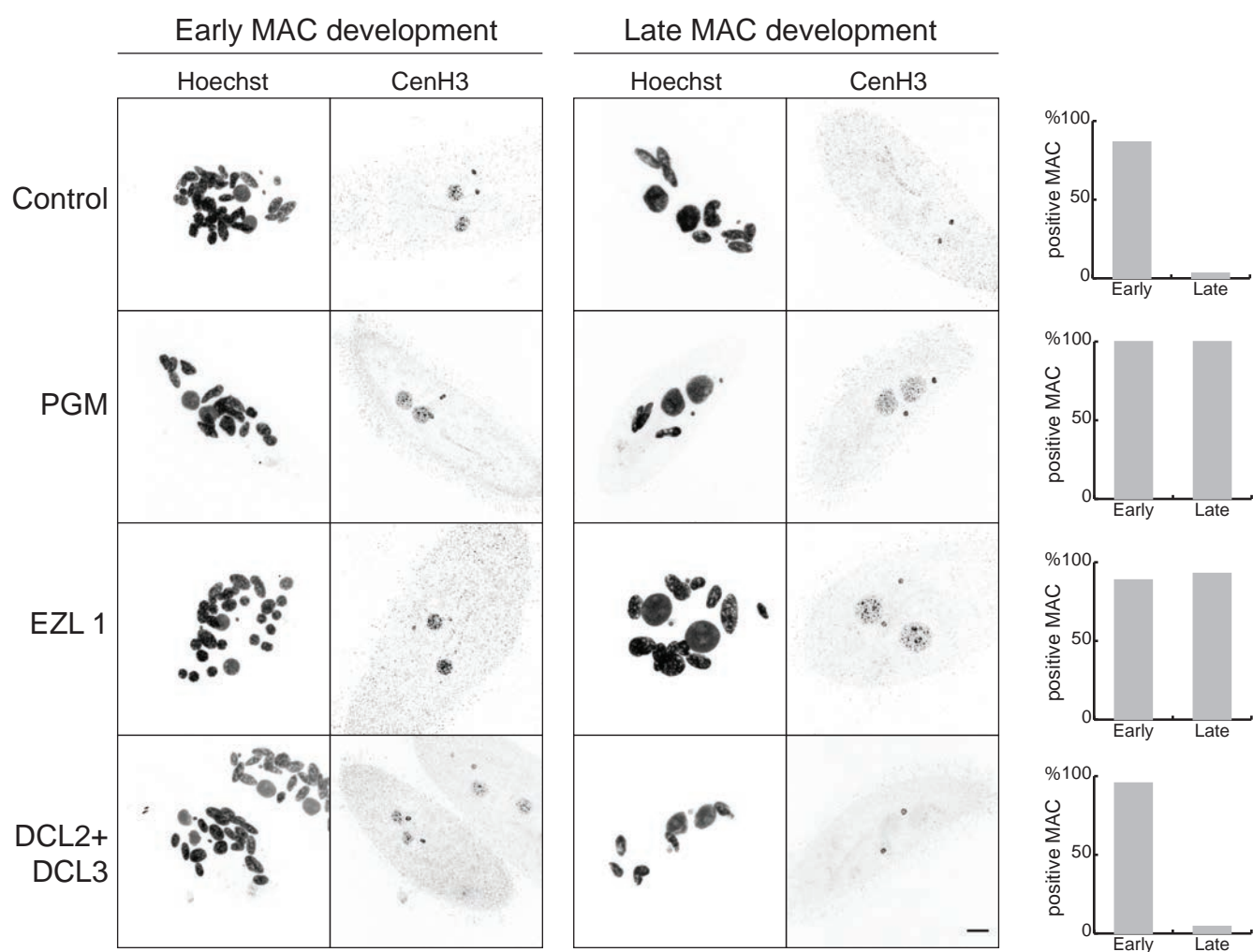


Figure 6. Role of the genes involved in DNA rearrangements in centromere inactivation.

Immunostaining with anti-CenH3 antibody at early and late stages of the macronuclear development on cells that have been knock down by RNAi for PiggyMac (PGM), Enhancer of zeste-like 1 (EZL1), Dicer-like 2 (DCL2) and Dicer-like3 (DCL3). Control cells have undergone an RNAi treatment with an empty vector. The scale bar represents 10 μ m. Diagrams on the right show, for one experiment, the quantification of cells containing CenH3 positive developing macronuclei at the two stages of the macronuclear development in each RNAi condition (same results have been reproduced in three independent experiments) number of counted cells > 100.

Table S1. Accession numbers of H3 and H3 variant genes of *Paramecium*

Name	Protein	Accession number
H3P1	H3	GSPATG00028569001 GSPATG00029116001 GSPATG00031217001 GSPATG00034503001 GSPATG00035465001
H3P1_2	H3	GSPATG00018242001
H3P3	H3	GSPATG00010278001 GSPATG00019420001 GSPATG00033453001
H3P2	H3.3	GSPATG00036420001
H3P4	H3.3	GSPATG00018066001
H3P5	H3 variant	GSPATG00027423001
H3M1	H3 variant	GSPATG00011562001
H3M2	H3 variant	GSPATG00011115001
CenH3-1	Centromeric variant	GSPATG00039439001
CenH3-2	Centromeric variant	GSPATG00020538001

LEGENDS

Figure 1. Multiple sequence alignments of *Paramecium* H3 proteins and H3 variants. The five canonical H3 proteins of *Paramecium tetraurelia*, H3P1, H3P12, H3P2, H3P3 and H3P4 are highly conserved between them and are conserved with the canonical human H3 histone (HsH3). Lysines K9 and K27 of the N-terminal tail are indicated in red. Five histone variants much more divergent have been found: H3P5, H3M1 and H3M2 have still unknown functions whereas CenH3-1 and CenH3-2 are two putative centromeric histone variants. CenH3-1 and CenH3-2 have very divergent N-terminal regions compared to canonical H3 proteins and contain additional amino acids in the first loop of the histone fold domain (indicated in red).

Figure 2. CenH3-1 and CenH3-2 expression profiles during *Paramecium* life cycle. RNA sequencing data (Arnaiz et al., *in preparation*) showing CenH3.1 (black line) and cenH3.2 (gray line) levels of expression in vegetative cells (V1.2 and V1), in starved cells (S1 and S1.2), during the sexual cycle and the macronuclear development (T0 to T75 hours). Some time points are duplicated because RNAs coming from two independent developmental time courses have been sequenced. FPKM: fragments per kilobase per million

Figure 3. CenH3-1 localizes in micronuclei during the entire cell cycle of *Paramecium*. **A.** Immunostaining with a polyclonal anti-CenH3 antibody on *Paramecium* cells (a) in interphase, (b) metaphase, (c) anaphase and (d) during cell division. Pictures are z projections of several confocal sections. Dashed lines indicate the shape of cells. In the merge panel, red color corresponds to the hoechst fluorescence and green color corresponds to the CenH3 fluorescence. The scale bar indicates 10 μ m. **B.** Confocal section of an interphasic *Paramecium* cell expressing a CenH3-GFP fusion protein. CenH3-GFP localizes only in micronuclei.

Figure 4. Centromere inactivation occurs during macronuclear development. A. Immunostaining with anti-CenH3 antibody on *Paramecium* conjugating cells allowing the observation of CenH3 localization during early sexual events. (a) Prophase of meiosis I. (b) Anaphase of meiosis I. (c) End of meiosis II. Eight haploid nuclei are present in cells. (d) karyogamy, CenH3 is present in the zygotic nucleus but has

disappeared from the seven degenerating haploid nuclei. (e) First division of the zygotic nucleus. **B.** Immunostaining with anti-CenH3 antibody on autogamous *Paramecium* cells. (a) Second division of the zygotic nucleus. (b-c) Early macronuclear development. CenH3 is present in the new macronuclei and in the new micronuclei. (d) Late macronuclear development. CenH3 has disappeared from the developing macronuclei. (e) Karyonide. (c-d) Red arrows indicate the developing macronuclei. (a-e) Dashed lines represent the shape of the cells. The scale bar indicates 10 μ m.

Figure 5. Centromere inactivation is concomitant with DNA elimination related events. **A.** Cytological evolution of a cell population in which the sexual cycle has been triggered by starvation. Time zero (0 hr) corresponds to the first observation of meiotic cells (about 35% of meiotic cells). Following time points have been chosen arbitrarily until 80 hours. **B.** Proportion of cells containing two CenH3 positive developing macronuclei during the developmental time course experiment presented in A. Centromere inactivation occurs between 20 hrs and 35 hrs. **C.** DNA rearrangements analysis in the developmental time course experiment presented in A. PiggyMac and T1b levels of expression have been followed by RT-PCR. Circular excised products of the IES 51A4578 have been detected by reverse PCR and appear to be actively degraded between 20 and 35 hours. Maternal inheritance of macronuclear deletions of the ND7 gene has been monitored by PCR. Elimination of the micronuclear forms in the developing macronuclei occurs between 20 hrs and 35 hrs.

Figure 6. Role of the genes involved in DNA rearrangements in centromere inactivation. Immunostaining with anti-CenH3 antibody at early and late stages of the macronuclear development on cells that have been knock down by RNAi for PiggyMac (PGM), Enhancer of zeste-like 1 (EZL1), Dicer-like 2 (DCL2) and Dicer-like3 (DCL3). Control cells have undergone an RNAi treatment with an empty vector. The scale bar represents 10 μ m. Diagrams on the right show, for one experiment, the quantification of cells containing CenH3 positive developing macronuclei at the two stages of the macronuclear development in each RNAi condition (same results have been reproduced in three independent experiments). Number of counted cells > 100.

METHODS

Strains and growth conditions

Experiments have been realized with *Paramecium tetraurelia* strain 51 wild-type. Cells were grown at 27°C in a wheat grass powder infusion (WGP, Pines International Inc.), inoculated with *Klebsiella Pneumoniae* bacteria and supplemented with 0.8 mg/mL β -sitosterol (Merck). To perform conjugation, starvation was induced in cells of each mating types to trigger cell reactivity. Autogamy was induced by starvation of about 30 divisions cells of unique mating types. For the developmental time course experiment, we used the variant strain 51 Δ ND7 Δ A carrying heritable macronuclear deletions of the genes *ND7* and *A* but harboring wild-type micronuclei (Garnier et al., 2004). Progression of the developmental stages has been monitored by nuclei hoechst staining.

Sequences alignments

Canonical H3 genes and H3 variant genes were identified in the macronuclear genome of *P. tetraurelia* by BLAST searches with the human H3 gene at ParameciumDB (<http://paramecium.cgm.cnrs-gif.fr/>) (Arnaiz et al., 2007). Corresponding protein sequences were aligned using the CLUSTAL Omega (1.1.0) software.

RNAi-mediated gene inactivation

RNAi Plasmids. CenH3-1 and CenH3-2 RNAi plasmids are derivatives of vector L4440 (Timmons and Fire, 1998) and carry a specific gene fragment between two convergent T7 promoters. For CenH3.1, we used a 168 bp fragment between position 7 and 175 of accession number GSPATT00039439001 and for CenH3.2, we used a 176 bp fragment between positions 7 and 182 of accession number GSPATT00020538001. PGM, EZL1, DCL2 and DCL3 plasmids are described respectively in (Baudry et al., 2009) ; (Lhuillier-Akakpo et al. *in preparation*) and (Lepère et al., 2009).

Preparation of silencing medium. dsRNA-producing plasmids were transformed into *E. coli* HT115 (Timmons et al., 2001). Silencing media were prepared basically as described in (Galvani and Sperling, 2002), by diluting precultures of the appropriate bacterial strains into WGP medium containing 0.1 mg/mL ampicillin. Following 6–8 h

of shaking at 37°C, bacterial cultures were diluted into the same medium to reach an OD₆₀₀ of 0,08 and IPTG was added (0.4 mM final) to induce the synthesis of dsRNA. Following overnight induction at 37°C, silencing media were supplemented with 0.8 mg/mL β -sitosterol before use.

RNAi during vegetative growth. For CenH3 genes silencing, *P. tetraurelia* cells were grown in standard *K. pneumoniae* medium for 20–30 vegetative divisions. Then, one cell was transferred each day in 200 μ L of freshly prepared silencing medium for one to five days of vegetative divisions. Cell growth was monitored by counting the number of cell divisions each day. To test micronuclei integrity, cells that had undergone one to three days of silencing were reintroduced in 800 μ L of normal medium and starved to trigger the sexual cycle. 12-48 post-autogamous cells were transferred individually in 200 μ L of *K. pneumoniae* medium to monitor growth of sexual progeny.

RNAi during autogamy. For CenH3-1 and CenH3-2 silencing, *P. tetraurelia* cells were grown in standard *K. pneumoniae* medium for 20–30 vegetative divisions then washed twice in silencing medium. About 50 exponentially growing cells were starved in 400 μ L of each silencing medium. Autogamy was monitored by hoechst staining, and cells were generally 100% autogamous at day 1 of starvation After 3 days, 12 to 48 cells were picked and transferred individually to 200 μ L of *K. pneumoniae* medium to monitor growth of sexual progeny. PiggyMac, EZL1, DCL2 and DCL3 silencing were performed identically except that cells were grown for two days of vegetative fissions in silencing medium (freshly induced medium was added the second day) before starvation-induced autogamy.

CenH3-GFP fusion construction

A fragment containing the CenH3-1 coding sequence was inserted into the plasmid pTI (Marker et al., 2010) upstream the GFP coding sequence adapted to *Paramecium* codon usage (Nowacki et al., 2005) with a 3' linker sequence (TCTAGGCCTGTTGCTACTGCA).

Immunofluorescences

About 200 cells were permeabilized for 5 min at RT in 100 μ L of PHEM-Triton (60 mM PIPES, 25 mM HEPES, 10 mM EGTA, 2 mM MgCl₂, 1% Triton, pH 6.9) and

fixed by addition of 200 μ L of formaldehyde 2% diluted in PHEM 1X for 30 min. Cells were washed twice in TBST-BSA (10 mM Tris pH 7.4, 150 mM NaCl, 0.1% Tween 20, 10 mM EGTA, 2 mM $MgCl_2$, 3% BSA) and incubated for 10 min at RT in 150 μ L of primary antibody diluted 1:500 in TBST-BSA. After two washings of 10 min in TBST-BSA, cells were incubated in 100 μ L of secondary antibody (Alexa Fluor A448 conjugated goat anti-rabbit IgG, Invitrogen), diluted 1:500 in TBST-3BSA for 10 min at RT. DNA was stained with hoechst (0,4 μ g/mL) and cells were mounted on slides in a drop of antifading reagent (Citifluor, AF2). Images have been acquired on a confocal microscope (Leica DMI 6000) with a 63x lens, z sections have been realized every 0,5 μ m on the entire depth of the cells.

RNA extraction and RT-PCR

Total RNAs were extracted from cells silenced for a control gene (*ICL7*) during the vegetative growth and at different stages of the sexual cycle. For each point of the developmental time course experiment, 200 mL of cells (about 800 000 cells) were centrifuged and frozen in liquid nitrogen for conservation. Total RNAs were extracted, reverse-transcribed and RT-PCR were performed as described in Lhuillier-Akakpo et al, *in preparation*.

DNA extraction and PCR

Total genomic DNA was extracted from 400 mL of cells (about 10^6 cells) at each time point of a developmental time course experiment in which a non-relevant gene (*ICL7*) has been silenced. DNA extraction, and PCR were performed with the same procedure than in Lhuillier-Akakpo et al, *in preparation*.

REFERENCES

- Arnaiz, O., Cain, S., Cohen, J., and Sperling, L. (2007). ParameciumDB: a community resource that integrates the *Paramecium tetraurelia* genome sequence with genetic data. *Nucleic Acids Res.* 35, D439–D444.
- Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., et al. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444, 171–178.
- Baudry, C., Malinsky, S., Restituto, M., Kapusta, A., Rosa, S., Meyer, E., and Bétermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev.* 23, 2478–2483.
- Bétermier, M., Duhaucourt, S., Seitz, H., and Meyer, E. (2000). Timing of Developmentally Programmed Excision and Circularization of *Paramecium* Internal Eliminated Sequences. *Mol. Cell. Biol.* 20, 1553.
- Blower, M.D., and Karpen, G.H. (2001). The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nat. Cell Biol.* 3, 730–739.
- Galvani, A., and Sperling, L. (2002). RNA interference by feeding in *Paramecium*. *Trends Genet.* 18, 11–12.
- Garnier, O., Serrano, V., Duhaucourt, S., and Meyer, E. (2004). RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. *Mol. Cell. Biol.* 24, 7370–7379.
- Hake, S.B., and Allis, C.D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: The “H3 barcode hypothesis.” *Proc. Natl. Acad. Sci.* 103, 6428–6435.
- Lepère, G., Nowacki, M., Serrano, V., Gout, J.-F., Guglielmi, G., Duhaucourt, S., and Meyer, E. (2009). Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res.* 37, 903–915.
- Marker, S., Mouël, A.L., Meyer, E., and Simon, M. (2010). Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in *Paramecium tetraurelia*. *Nucleic Acids Res.* 38, 4092–4107.
- Nowacki, M., Zagorski-Ostojka, W., and Meyer, E. (2005). Nowa1p and Nowa2p: Novel Putative RNA Binding Proteins Involved in trans-Nuclear Crosstalk in *Paramecium tetraurelia*. *Curr. Biol.* 15, 1616–1628.
- Stoler, S., Keith, K.C., Curnick, K.E., and Fitzgerald-Hayes, M. (1995). A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. *Genes Dev.* 9, 573–586.
- Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. *Nature* 395, 854–854.

Timmons, L., Court, D.L., and Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* 263, 103–112.

RESULTATS COMPLEMENTAIRES

Identification de séquences centromériques par immunoprécipitation de la chromatine (ChIP)

Nous avons montré que l'inactivation des centromères au cours du développement macronucléaire dépend de gènes impliqués dans les réarrangements du génome. Ces résultats suggèrent fortement que les séquences centromériques sont physiquement éliminées du génome macronucléaire. Pour valider cette hypothèse, nous devons identifier des séquences d'ADN centromériques chez *P. tetraurelia* et montrer que ces dernières sont absentes du génome macronucléaire. Nous montrerions alors pour la première fois que des séquences mic-spécifiques assurent une fonction micronucléaire essentielle. L'identification de séquences centromériques apporterait également des informations importantes sur la nature des centromères chez cet organisme, d'autant qu'aucune région centromérique n'est à ce jour connue dans le phylum des ciliés. Les centromères sont-ils constitués de séquences répétées de type satellite, d'éléments transposables ou encore de séquences uniques ? Que peut-on apprendre de ces séquences quant à l'évolution des régions centromériques chez les eucaryotes ? La seule approche possible nous permettant d'identifier de l'ADN centromérique consiste à aller pêcher, par immunoprécipitation, l'ADN physiquement associé à CenH3 et à le séquencer. Pour cela, j'ai d'abord dû établir les conditions expérimentales adaptées à la réalisation du ChIP chez *Paramecium*. Par la suite, cette technique pourra également être appliquée à d'autres cibles comme la triméthylation de H3K9 et H3K27 pour déterminer si ces marques sont associées aux séquences à éliminer dans les macronoyaux en développement (voir Résultats, article I en préparation). De manière générale, la mise au point du ChIP offrira à l'avenir un outil crucial au laboratoire pour étudier le rôle de facteurs impliqués dans l'organisation chromatiniennne pendant les réarrangements du génome.

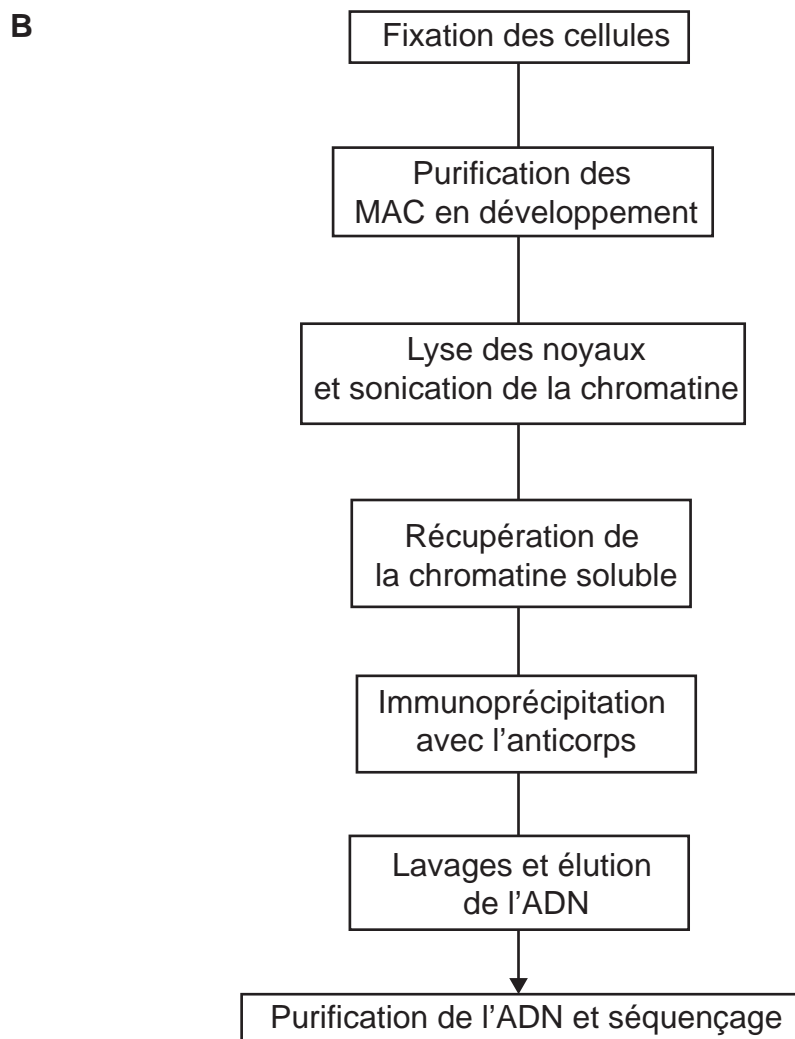
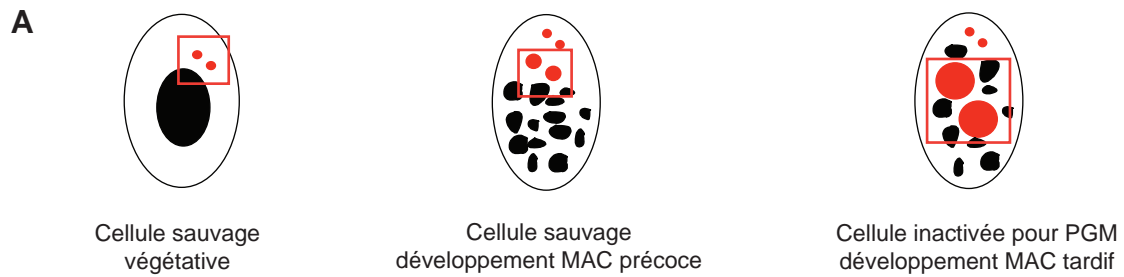


Figure 19. A. Représentation schématique des différentes localisations possibles de CenH3 dans les cellules de *Paramecia*. Les noyaux contenant la protéine CenH3 endogène sont colorés en rouge et les noyaux à utiliser pour le ChIP sont encadrés en rouge. **B.** Les différentes étapes du ChIP.

I. ChIP CenH3 : choix de la stratégie

1. Matériel de départ : mic ou MAC ?

Comme montré précédemment (Article 2 en préparation), dans les cellules végétatives, la protéine CenH3 est présente uniquement dans les micronoyaux. Une stratégie intéressante consisterait donc à immunoprécipiter la chromatine micronucléaire associée à CenH3 pour identifier les séquences qui constituent les centromères actifs de *Paramecium*. Cependant, les deux micronoyaux qui contiennent chacun deux copies du génome mic (environ 100 Mb) représentent à peine 0,7% de la chromatine cellulaire totale face au macronoyau qui contient 800 copies du génome MAC (72 Mb). Dans un lysat cellulaire, la chromatine mic serait donc « noyée » dans l'abondance de la chromatine MAC rendant la mise au point du ChIP particulièrement difficile. De manière encourageante, la purification de micronoyaux a récemment été rendue possible au laboratoire par centrifugation différentielle de lysats cellulaires associée à des techniques de tri cellulaire (FACS) utilisant la protéine de fusion CenH3-GFP, marqueur des micronoyaux (F. Guérin, communication personnelle). Cependant, les rendements obtenus sont encore très faibles et difficilement compatibles avec les techniques de ChIP. De manière intéressante, la présence de CenH3 dans les macronoyaux en développement dans les stades précoces du développement offre la possibilité d'une approche alternative se basant sur de la chromatine macronucléaire, plus abondante. De plus, dans les conditions particulières d'inactivation de PiggyMac ou de Ezl1, CenH3 est également maintenue dans les macronoyaux jusque dans les stades tardifs du développement (Figure 19A). J'ai donc tiré avantage de ces situations pour mettre au point des conditions de ChIP avec de la chromatine macronucléaire.

2. Protéine cible : CenH3 ou CenH3-tag ?

Comme les protéines CenH3 sont très divergentes entre organismes, nous avons fait produire un anticorps polyclonal de lapin dirigé spécifiquement contre la protéine CenH3.1 de *Paramecium tetraurelia* en utilisant un peptide spécifique de la région N-terminale de la protéine. Les capacités de cet anticorps à reconnaître sa protéine cible ont été prouvées en immunofluorescence mais nous ne connaissons pas son pouvoir immunoprécipitant et de manière inquiétante, il n'a pas été possible à ce jour de détecter CenH3.1 par western blot avec cet anticorps. Il paraît donc

judicieux de tenter des premières expériences de ChIP avec un anticorps générique commercial, connu pour être efficace en ChIP et dirigé contre une étiquette (GFP, Flag...) en utilisant des cellules qui auraient été transformées avec la protéine de fusion correspondante (CenH3-GFP, CenH3-Flag...). Cependant, au cours du développement macronucléaire, l'expression des transgènes, qui est maternelle, est progressivement supplantée par l'expression zygotique des gènes endogènes correspondant et le maintien d'une protéine de fusion dans la cellule en développement dépend de la rapidité de son turn-over. Concernant CenH3-GFP, nous avons observé que le turn-over de cette protéine était très court et que le signal de la GFP disparaissait très précocement dans les macronoyaux en développement, très probablement avant le moment de la disparition de CenH3 endogène. Comme il existe obligatoirement une asynchronie de 5 à 6 heures dans une population de cellules en développement, il devient alors très difficile de produire une population homogène de cellules contenant toutes la protéine de fusion CenH3-GFP dans leurs macronoyaux en développement. De même, nous avons observé que l'inactivation de PiggyMac, ne permettait pas la rétention de CenH3-GFP à la fin du développement macronucléaire, alors que la protéine endogène est maintenue, car l'expression maternelle a été stoppée. Par conséquent, l'utilisation d'une étiquette n'est pas compatible avec la stratégie développée en (1) consistant à utiliser les macronoyaux en développement comme source de chromatine. J'ai donc utilisé l'anticorps anti-CenH3 pour cibler la protéine endogène en espérant qu'il montre de bons résultats.

3. ChIP natif ou crosslinking de la chromatine?

Les histones étant naturellement fortement associées à l'ADN, les expériences de ChIP dirigées contre des protéines histones sont très souvent réalisées en conditions natives, c'est-à-dire sans traitement chimique permettant la formation de liaisons covalentes ADN-protéines dans la chromatine (crosslink). La chromatine est dans ce cas digérée en mononucléosomes par la nucléase de microcoque, une enzyme qui digère l'ADN préférentiellement dans les régions linker. Les conditions natives ont pour avantage de préserver les épitopes reconnus par les anticorps et limitent les résultats faux-positifs. A l'inverse, le crosslink de la chromatine est généralement requis pour immunoprécipiter la chromatine associée à des protéines non-histones, faiblement associées à l'ADN mais peut altérer les épitopes et générer des associations ADN-protéines ectopiques produisant des résultats faux-positifs. Bien que CenH3 soit une histone, j'ai choisi de

commencer par faire des expériences de ChIP en crosslinkant la chromatine en espérant augmenter les chances de résultats positifs avec l'anticorps anti-CenH3.

II. Mise au point du ChIP chez *Paramecium tetraurelia*

Parce qu'aucun protocole de ChIP n'était établi chez *Paramecium*, je me suis inspiré de nombreux protocoles développés chez divers organismes (cellules humaines, plantes, *Tetrahymena*...) pour tenter de trouver les conditions adéquates à la réalisation d'expériences de ChIP chez *Paramecium* avec l'anticorps anti-CenH3. Les différentes étapes du ChIP sont rappelées dans la Figure 19B.

1. Préparation et sonication de la chromatine macronucléaire

Fixation des cellules : pour minimiser les chances d'altérer l'interaction épitope-anticorps par la fixation, j'ai utilisé des conditions de fixation similaires à celles utilisées pour les expériences d'immunofluorescence que l'on sait être favorables à la reconnaissance de CenH3 par l'anticorps: les cellules sont fixées avec 1% de formaldéhyde (Sigma) dans du PHEM 1X (Pipes 60 mM, Hepes 25 mM, EGTA 10 mM, MgCl₂ 2 mM, pH 6,9) pendant 10 min. La réaction est ensuite arrêtée avec de la glycine à 125 mM final pendant 5 min puis les cellules sont lavées deux fois dans du Tris 10 mM pH 7,4. Les cellules fixées peuvent être congelées à -80°C à cette étape ou utilisées pour l'étape suivante.

Purification des MAC en développement : la purification des macronoyaux est une étape indispensable pour préserver la qualité de la chromatine. En effet, j'ai pu observer que la chromatine soluble issue de cellules entièrement lysées était fortement dégradée après une nuit à 4°C, très certainement parce qu'il existe dans le cytoplasme des nucléases qui, même après traitement à la formaldéhyde dégradent l'ADN cellulaire quand les noyaux sont lysés (Figure 20A). Il est pourtant nécessaire de maintenir l'intégrité de la chromatine soluble plusieurs heures à 4°C pour pouvoir réaliser les étapes suivantes du ChIP. Le problème peut être résolu en purifiant les macronoyaux par centrifugation différentielle ce qui permet d'éliminer la fraction cytoplasmique riche en enzyme de dégradation (protocole décrit dans Arnaiz et al., 2012). Après purification des macronoyaux, les fragments d'ADN incubés une nuit à 4°C ont un poids

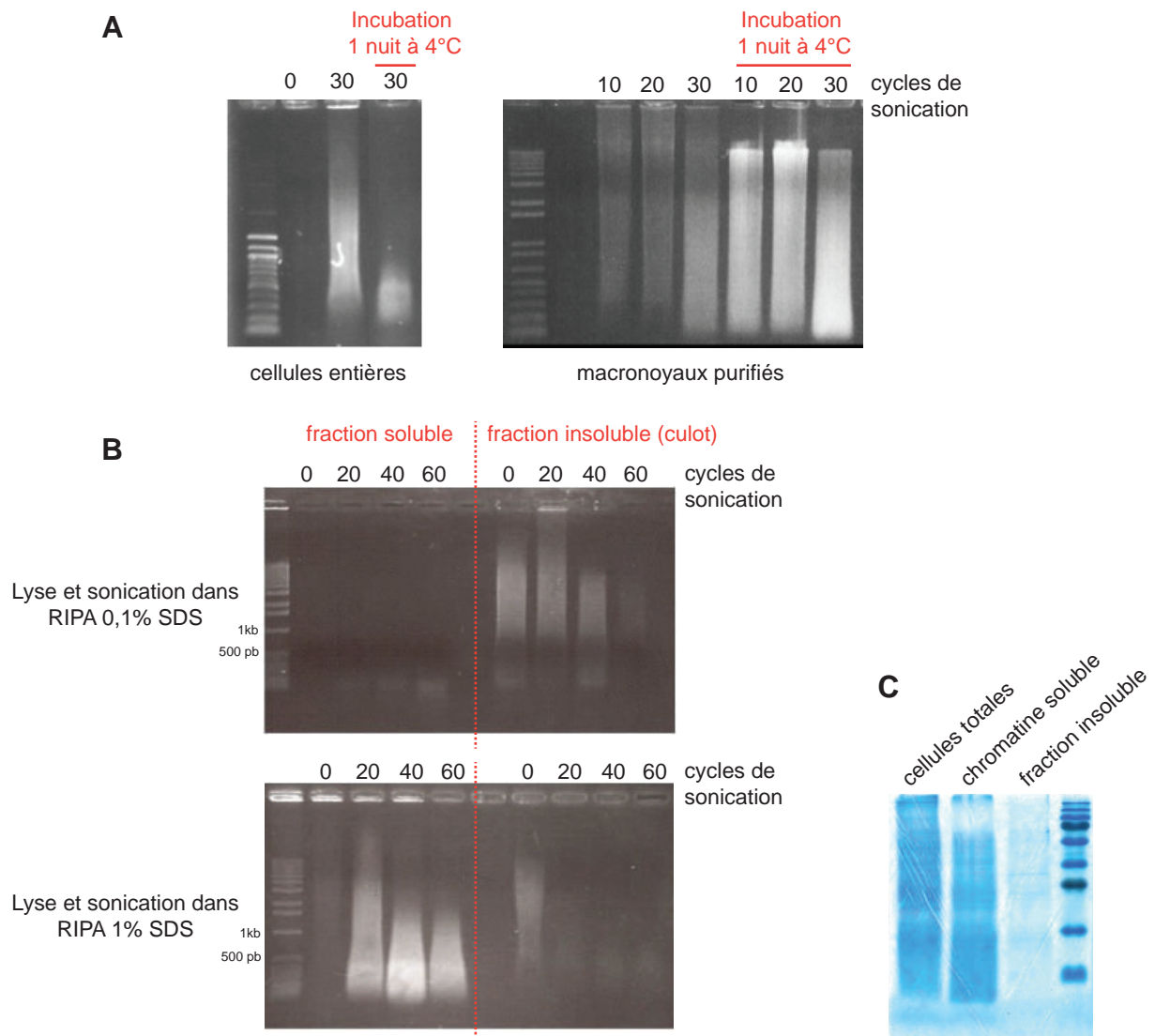


Figure 20. A. Comparaison de la taille des fragments d'ADN issus de la chromatine macronucléaire avant et après incubation de cette chromatine une nuit à 4°C. Sur le gel d'agarose de gauche, la chromatine est issue de cellules entières lysées, sur le deuxième gel les macronoyaux ont été purifiés avant d'être lysés. **B.** Comparaison du contenu en ADN de la fraction soluble (surnageant) et de la fraction insoluble (culot) dans le cas où les noyaux ont été soniqués en présence de 0,1% de SDS (gel supérieur) ou de 1% de SDS (gel inférieur). Il faut 1% de SDS pour que la chromatine soit dans la fraction soluble. **C.** Gel d'acrylamide montrant les protéines totales issues de cellules entières et des fractions soluble et insoluble après sonication de la chromatine.

moléculaire équivalent à celui avant incubation (Figure 20A). Une fois purifié, les macronoyaux purifiés peuvent être congelés à -80°C ou utilisés pour l'étape suivante.

Lyse des noyaux et sonication de la chromatine : Pour préparer la chromatine soluble qui servira aux expériences de ChIP, les noyaux sont lysés dans un tampon RIPA (Tris 50 mM pH 8, NaCl 150 mM, EDTA 10 mM, NP-40 1%, deoxycholate de sodium 0,5%, supplémenté extemporanément en SDS 1%, PMSF 1 mM et inhibiteurs de protéases 1X) et la chromatine est soniquée pour obtenir des fragments d'ADN de taille comprise entre 500 et 1000 pb. Après plusieurs tests, j'ai pu conclure qu'il fallait 40 cycles de sonication (15 secondes ON/15 secondes OFF), soit 20 min de sonication pour obtenir des fragments de taille acceptable à partir de 200 µL de cellules entières à 10⁶ cellules/mL (Figure 20B, panel inférieur), comme à partir de 200 µL de noyaux purifiés à 10⁶ MAC/mL (non montré). La chromatine est ensuite centrifugée (10 min à 4°C à 16 000 g) pour éliminer les débris membranaires et récupérer la chromatine soluble dans le surnageant. Bien que le SDS puisse être délétère pour les anticorps dans les étapes suivantes du ChIP, il ne faut pas trop diminuer la concentration de SDS pour la lyse des noyaux et la sonication car il semble que le SDS favorise l'extraction de la chromatine soluble. En effet, avec seulement 0,1% de SDS dans le tampon RIPA, toute la chromatine reste dans le culot après centrifugation (fraction insoluble) et ne peut donc pas être utilisée pour le ChIP (Figure 20B, panel supérieur). La lyse et la sonication sont donc réalisées avec 1% de SDS mais la chromatine sera diluée dix fois pour l'immunoprécipitation et la concentration finale de SDS sera abaissée à 0,1%. Pour vérifier que les conditions de lyse et de sonication n'ont pas altéré les protéines et permettent leur bonne extraction, j'ai également observé l'enrichissement des protéines totales dans la chromatine soluble par rapport à la fraction insoluble sur un gel d'acrylamide. (Figure 20C).

2. Immunoprécipitation

Les conditions d'immunoprécipitation de la chromatine étant moins dépendante de l'organisme modèle, j'ai choisi d'utiliser un kit commercial (société Ademtech) pour réaliser les étapes suivantes d'immunoprécipitation. Ce kit contient des billes magnétiques couplées à la protéine A permettant une forte interaction avec les IgG de lapin (isotype de l'anticorps CenH3), des solutions de lavages et une solution d'élution. Afin d'éliminer une partie des interactions aspécifiques entre les billes magnétiques et la chromatine, celle-ci est incubée pendant une heure à

4°C avec les billes seulement pour une étape de « pre-clearing » avant d'être incubée pendant la nuit à 4°C avec les billes magnétiques couplées à l'anticorps d'intérêt. La quantité de chromatine à utiliser pour chaque réaction de ChIP doit représenter 20 à 30 ug d'ADN ce qui correspond dans notre cas à 50-75 uL d'une solution de noyaux à 10^6 MAC/mL. Il est difficile de déterminer la quantité adéquate d'anticorps à utiliser, celle-ci étant variable d'un anticorps à l'autre et souvent déterminée empiriquement mais une quantité de 1 à 10 µg est généralement préconisée. Les lavages et l'élution de l'ADN sont réalisés suivant les informations du kit Ademtech. L'éluat est finalement incubé à 65°C pendant une nuit avec de la protéinase K pour la réversion du crosslink et la dégradation des protéines ce qui permettra l'extraction de l'ADN immunoprécipité.

3. Test qualité

La première difficulté dans la mise au point du ChIP chez *Paramecium* est l'absence de contrôle positif qui permettrait de valider que les conditions utilisées sont adéquates. Par exemple, connaît-on des protéines comme des facteurs de transcription qui se fixeraient spécifiquement à certaines séquences d'ADN dont nous pourrions valider l'interaction par ChIP? Pas vraiment... Il est cependant connu, par exemple, que la triméthylation de la lysine 4 sur l'histone H3 (H3K4me3) est associée aux séquences promotrices chez les eucaryotes (Bernstein et al., 2002; Heintzman et al., 2007) Cette région de l'histone H3 est très conservée entre l'Homme et *Paramecium* ce qui pourrait nous permettre d'utiliser des anticorps commerciaux validés en ChIP pour révéler les régions promotrices du génome macronucléaire et nous servir de contrôle positif pour valider les conditions de ChIP. Cependant, les tentatives d'immunofluorescence avec un anticorps commercial dirigé contre H3K4me3 (abcam, ab8580) ne nous ont donné aucune évidence de la présence de cette marque dans les macronoyaux. La deuxième difficulté, qui est elle associée à l'anticorps anti-CenH3, est l'impossibilité de faire des western blot pour vérifier que CenH3 est bien présente à chaque étape du ChIP, qu'elle est notamment présente dans la chromatine soluble, qu'elle n'est pas dégradée par la sonication ou par certains tampons inadaptés et qu'elle est enrichie dans la fraction immunoprécipitée.

Pour tout de même tester les conditions de ChIP précédemment établies, j'ai essayé d'immunoprécipiter l'histone H4 penta-acétylé, une histone modifiée couvrant certainement une grande partie du génome macronucléaire et pour laquelle nous avons un anticorps commercial (Millipore, 06-946) validé en ChIP qui montre un signal très intense dans les macronoyaux en

immunofluorescence. J'ai pu montrer par western blot qu'après immunoprécipitation avec l'anticorps H4 penta-acétylé, l'histone a été enrichie de manière plus importante que dans la condition contrôle sans anticorps (Figure 21A). J'ai également réalisé une expérience de ChIP dirigée contre l'histone canonique H3 (Millipore, 07-690) après laquelle j'ai tenté d'amplifier par PCR une région macronucléaire choisie arbitrairement. De manière intéressante j'ai pu observer que de l'ADN avait été immunoprécipité dans les conditions de ChIP mais pas dans le contrôle sans anticorps (Figure 21B) et que l'enrichissement en ADN était d'autant plus important que la quantité d'anticorps utilisée était grande, ce qui est attendu pour une protéine comme H3 qui n'est pas limitante dans la chromatine (non montré).

Bien que les mêmes approches ne soient pas possible pour tester le ChIP avec l'anticorps CenH3 car celui-ci ne fonctionne pas en western blot et parce que nous ne connaissons aucune séquence d'ADN associée à CenH3, ces résultats encourageants m'ont conduite à réaliser des expériences de ChIP avec l'anticorps anti-CenH3 destinées au séquençage de l'ADN immunoprécipité.

III. ChIP-Seq : identification des séquences associées CenH3

Expérience 1

ChIP avec l'anticorps anti-CenH3 sur des macronoyaux en développement de cellules inactivées pour PiggyMac

La première expérience a été réalisée à partir d'une population de $1,8.10^6$ cellules issues de quatre clones indépendants (population polycaryonidale), inactivées pour PiggyMac par ARN interférence, 100% autogames, toutes bloquées au stade de deux macronoyaux en développement positifs pour CenH3 en immunofluorescence, environ 70 heures après le début de l'autogamie.

Les cellules ont été centrifugées puis fixées dans 3,6 mL de formaldéhyde diluée à 1% dans du PHEM 1X pendant 10 min à RT (température ambiante).

La réaction a été stoppée avec 225 μ L de glycine 2M (125 mM final) pendant 5 min à RT.

Après une centrifugation à 3000 rpm pendant 5 min à 4°C, les cellules ont été lavées 2 fois dans 10 mL de Tris 10mM pH 7,4 froid et le culot de cellules a été congelé à -80°C.

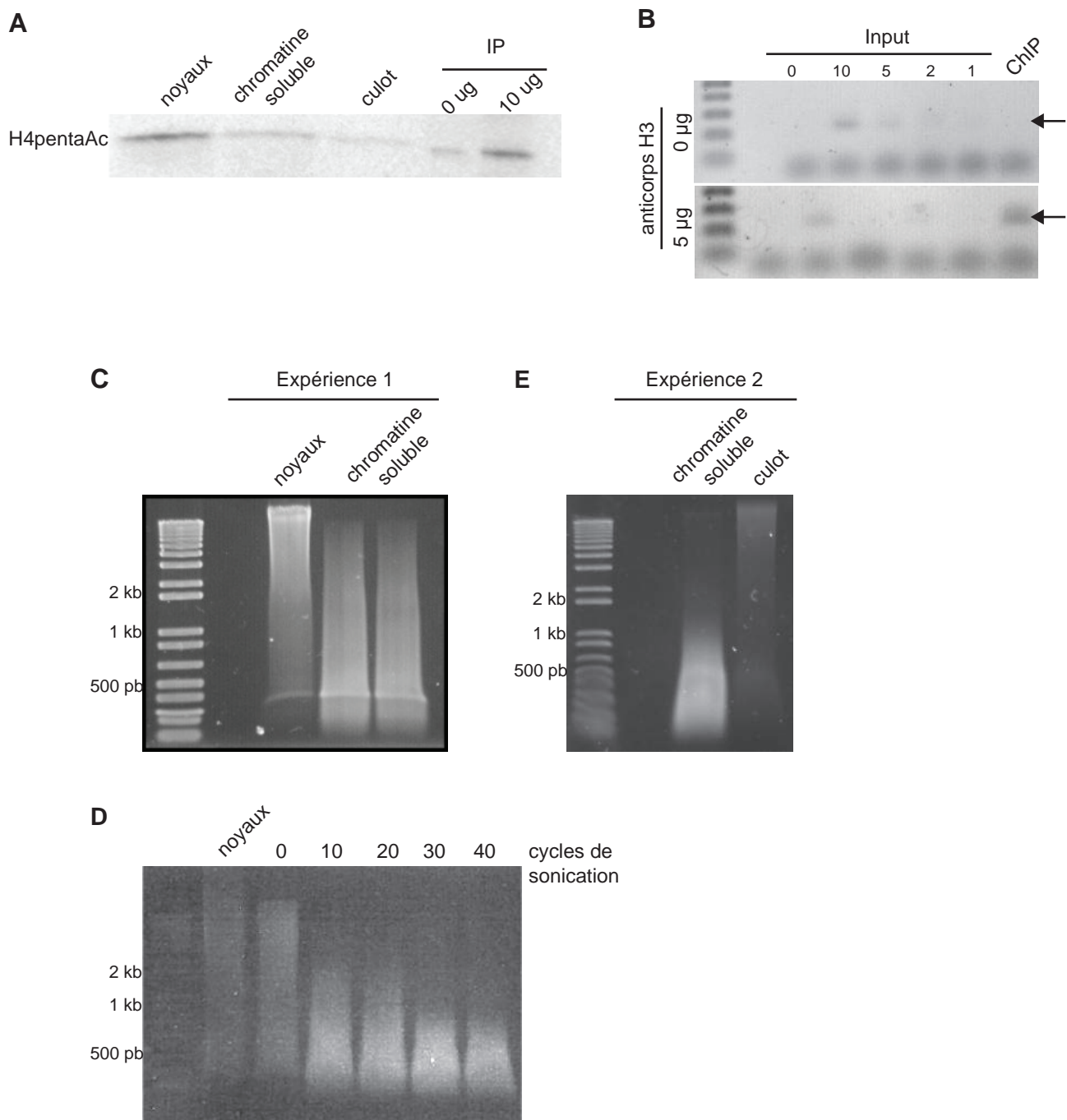


Figure 21. A. Western blot avec l'anticorps H4 penta-acétylé (H4pentaAc). Le même volume de protéines totales issues des macronoyaux purifiés avant sonication, de la chromatine soluble et de la fraction insoluble a été déposé. Ce même volume de chromatine soluble a été utilisé pour immunoprécipiter H4pentaAc avec 0 ou 10 µg d'anticorps et l'éluat protéique a été déposé. **B.** PCR dans une région macronucléaire réalisées sur différentes quantités d'ADN issu de la chromatine soluble (Input) et sur l'ADN issu du ChIP avec 0 ou 5 µg d'anticorps H3. **C.** Gel d'agarose 1% présentant l'ADN issu des noyaux avant sonication et de la chromatine soluble dans l'expérience 1. **D.** Taille des fragments d'ADN issus de noyaux fixés avec 0,5% de formaldéhyde après 0, 10, 20, 30 et 40 cycles de sonication (15 secondes ON/ 15 secondes OFF). **E.** Gel d'agarose 1% présentant l'ADN issu de la chromatine soluble et de la fraction insoluble dans l'expérience 2.

Quelques jours plus tard, les cellules ont été décongelées pour purifier les macronoyaux. 600 μ L de solution de lyse (3 volumes du culot de cellules initial) ont été ajoutés aux cellules décongelées. Après 5 min d'attente dans la glace, les cellules ont été cassées mécaniquement au potter (compter environ 100 aller-retour). Le lysat a été centrifugé à 1000 g, 1 min à 4°C et le culot de noyaux a été lavé deux fois avec 2 mL de solution de lavage (10 volumes du culot de cellules initial). Les noyaux ont été récupérés dans 500 μ L de coussin de sucrose et ont été centrifugés à 21000 g pendant 1 heure à 4°C sur 3 mL de coussin de sucrose.

Les macronoyaux ont été repris dans du tampon de lavage et comptés sur une cellule de Malassez. Après une centrifugation à 1000 g 1 min à 4°C, les noyaux ont été re-suspendus dans du tampon RIPA sans SDS à environ 10^6 MAC/mL et congelés par aliquots de 200 μ L à -80°C (congélation rapide dans un mélange de carboglace d'éthanol 100%).

Après décongélation, les aliquots ont été supplémentés SDS (1%) final, PMSF (1 mM final) et inhibiteurs de protéase (1X final). Les échantillons ont été vortexés pour aider la lyse des noyaux avant d'être soniqués pendant 40 cycles de 15 secondes ON / 15 secondes OFF en remplaçant l'eau du bain tous les 10 cycles pour garder une température autour de 6-7°C. La Figure 21C montre une image des fragments d'ADN séparés par électrophorèse avant et après sonication.

Pendant ce temps, les billes magnétiques (25 μ L par condition) sont incubées avec un agent bloquant puis avec 2 μ g ou 4,5 μ g d'anticorps CenH3 et avec 4,5 μ g de sérum pré-immun du lapin qui a servi à produire l'anticorps CenH3 (contrôle négatif).

Les échantillons de chromatine soniquée ont été centrifugés et la chromatine soluble récupérée. Une partie de la chromatine soluble est détournée pour en extraire l'ADN et constituera la fraction « Input ». Le reste de la chromatine soluble est diluée 10 fois dans du tampon RIPA sans SDS (75 μ L dans 675 μ L, la concentration finale de SDS est donc de 0,1%) et répartie dans autant de tubes que de conditions pour l'étape de « pre-clearing » avec 25 μ L de billes magnétiques vierges, 1h30 à 4°C. La chromatine a ensuite été incubée en rotation avec 25 μ L de billes magnétiques couplées à l'anticorps ou aux IgG du sérum pré-immun pendant une nuit à 4°C.

Les billes ont été lavées et l'ADN élué selon le protocole du kit Ademtech. Les protéines ont été digérées avec l'ajout de protéinase K à 10mg/mL pendant 2 heures à 37°C (agitation à 300 rpm).

Enfin l'ADN a pu être séparé des billes magnétiques pour la réversion du crosslink une nuit à 65°C (agitation à 300 rpm).

L'ADN immunoprécipité dans les trois conditions (anticorps anti CenH3 2 µg, anticorps anti-CenH3 4,5 µg et serum pré-immun 4,5 µg) a ensuite été extrait au phénol et précipité avec 1 mL d'éthanol 100%, 30 µL d'acétate de sodium pH 5 et 1 µL de glycogène (35 mg/mL). Les échantillons ont été dosés avec le kit QBit (Invitrogen), les concentrations obtenues sont reportées dans la Table 1.

Expérience 2

ChIP avec l'anticorps anti-CenH3 sur des macronoyaux de cellules sauvages au stade 15 heures de développement

Une deuxième expérience a été réalisée à partir d'une population de $1,9 \cdot 10^6$ cellules issues d'un même clone cellulaire (population monocaryonidale), 100% autogames, environ 15 heures après le début des événements sexuels. En immunofluorescence, 100% des cellules contiennent deux macronoyaux en développement positifs pour CenH3.

Dans l'expérience 1, il m'est apparu qu'après la sonication, la taille des fragments d'ADN était assez dispersée et, même si la majorité des fragments étaient de taille comprise entre 500 et 1000 pb, une proportion importante d'entre eux avait un poids moléculaire supérieur à 1kb, ce qui peut rendre le ChIP moins efficace. Pour ne pas augmenter le temps de sonication, déjà long, et risquer d'altérer la chromatine, j'ai décidé de diminuer la fixation des cellules pour augmenter l'efficacité de la sonication. En fixant les cellules avec seulement 0,5% de formaldéhyde, j'ai pu observer que 30 cycles de sonication suffisaient pour réduire la taille des fragments d'ADN entre 300 et 1000 pb et que la taille de ces fragments était beaucoup moins dispersée (Figure 21D).

Dans la deuxième expérience, les cellules ont donc été centrifugées puis fixées dans 3,8 mL de formaldéhyde diluée à 0,5% dans du PHEM 1X pendant 10 min à RT.

De manière complètement similaire à ce qui est décrit pour l'expérience 1, les macronoyaux ont été purifiés, resuspendus dans du tampon RIPA sans SDS à la concentration de 10^6 MAC/mL puis aliquotés en échantillons de 200 µL.

La chromatine soluble a été soniquée (Figure 21E) et utilisée pour des expériences de ChIP avec 3 µg d'anticorps CenH3, 5 µg d'IgG de lapin commerciales non spécifiques (contrôle négatif) et 5 µg d'anticorps H3K9me3.

Les quantités d'ADN qui ont été récupérées après le ChIP sont reportées dans la Table 1.

Séquençage de l'ADN et analyse des données (en collaboration avec O. Arnaiz, CGM, Gif-sur-Yvette)

L'ADN immunoprécipité a été séquencé grâce à la technologie Mi-Seq de Illumina sur la plateforme de génomique de Gif-sur-Yvette (ImaGif).

Après élimination des lectures de mauvaise qualité, les séquences ont été alignées sur le génome macronucléaire (MAC) et sur le génome macronucléaire contenant les IES (MAC+IES). Le nombre de lectures qui s'alignent sur le génome MAC dans les expériences 1 et 2 est reporté dans la Table 2. Comme attendu, la couverture du génome de l'échantillon Input est homogène sur le génome (Figure 22A). A l'inverse, dans les conditions de ChIP, seules certaines régions du génome sont enrichies. De manière frappante, les pics d'enrichissement observés dans la condition de ChIP CenH3 sont très similaires aux pics observés dans la condition contrôle (sérum pré-immun ou IgG). De plus, ces pics sont corrélés à des régions particulièrement riches en GC comme s'il existait un biais en faveur de ces régions dans la procédure de ChIP et/ou dans la procédure de séquençage (Figure 22A).

Le logiciel MACS indique les régions du génome qui présentent une différence significative d'enrichissement entre la condition d'intérêt et la condition contrôle. 50 régions pour l'expérience 1 et 303 régions pour l'expérience 2, présentent des différences d'enrichissement significatives entre l'échantillon CenH3 et l'échantillon contrôle sur le génome MAC+IES. Observées individuellement, ces régions présentent cependant un enrichissement très faible dans la condition CenH3 et ressemblent à un signal de bruit de fond. Au final, aucune séquence macronucléaire et aucune IES ne ressort comme étant particulièrement associée à CenH3. Ceci n'est pas tellement surprenant dans la mesure où il existe une forte probabilité pour que l'ADN centromérique se trouve dans les régions mic-spécifiques.

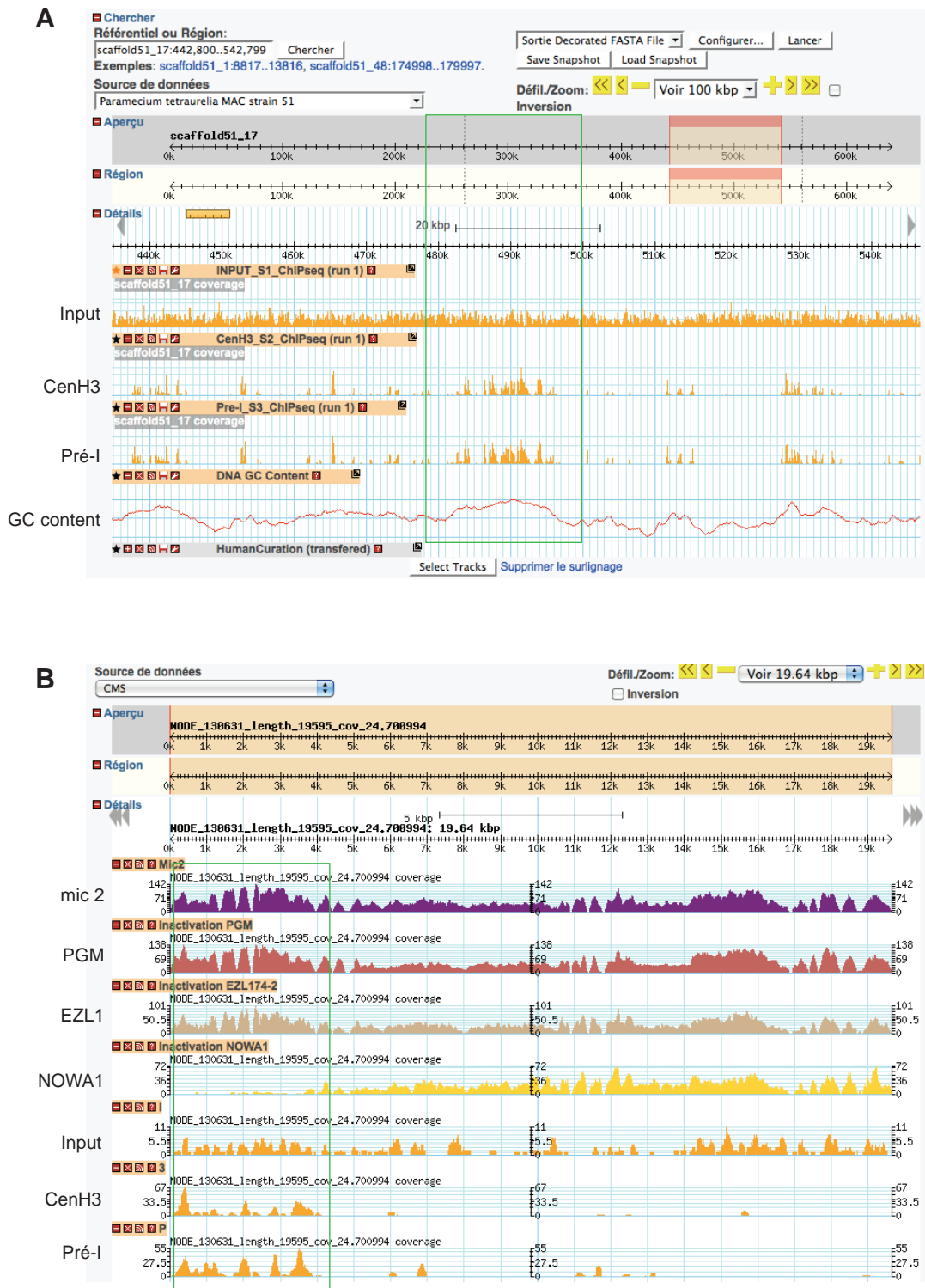


Figure 22. A. Vue de 100 kbp du scaffold 17 entre les positions 442,800 et 542,799. Alignement des séquences des échantillons Input (Input_S1_ChIPseq), ChIP CenH3 (CenH3_S2_ChIPseq) et contrôle (Pre-I_S3_ChIPseq) dans cette région. Comparaison avec le contenu en GC du génome (DNA GC content). Un exemple de région riche en GC et présentant un enrichissement dans les conditions de ChIP CenH3 et contrôle est encadré en vert. **B.** Vue de 19,64 kbp du contig micronucléaire NODE_130631. Une région couverte dans le génome micronucléaire (Mic2), dans le génome de cellules inactivées pour PiggyMac (PGM) et pour EZL1 mais absente du génome de cellules inactivées pour NOWA1 est encadrée en vert. La couverture de cette région dans les échantillons Input (I), ChIP CenH3 (3) et ChIP contrôle (P) est présentée en dessous.

Les séquences qui ne s'alignent pas sur le génome macronucléaire ou sur les IES ont donc été alignées sur des contig micronucléaires. Ces contig ont été obtenus récemment par l'assemblage de lectures issues du séquençage à haut débit de quelques nanogrammes d'ADN micronucléaire provenant de micronoyaux purifiés. 213 régions pour l'expérience 1 et 316 régions pour l'expérience 2, ont été identifiées par MACS comme significativement enrichies dans la condition CenH3. L'enrichissement réellement observé est une nouvelle fois très faible. Pour essayer de faire ressortir des zones intéressantes, d'autres traitements ont été appliqués :

1. La recherche de mots exceptionnels : des mots de 6, 15, 20, 25 ou 30 pb ont été recherchés dans les lectures des conditions CenH3 et contrôle dans l'idée que les centromères seraient des séquences répétées et contiendraient plus fréquemment certains mots.
2. L'assemblage *de novo* des lectures pour identifier des régions plus fortement couvertes dans l'échantillon CenH3 que dans l'échantillon contrôle.
3. Seules les régions d'ADN présentes dans le génome des cellules inactivées pour EZL1 et absentes du génome des cellules inactivées pour NOWA1 (même phénotype que DCL2 et DCL3 vis à vis de CenH3) sont prises en compte.

Quelques régions micronucléaires présentant un très léger enrichissement pour CenH3 et répondant à l'ensemble de ces critères ont été identifiées. Certaines, comme celle présentée sur la Figure 22B, se trouvent à l'extrémité d'un contig micronucléaire suggérant que des séquences d'intérêt pourraient se trouver à l'extérieur, dans des régions non assemblées du génome micronucléaire. L'analyse du séquençage des échantillons de l'expérience 2 déterminera si ces résultats sont reproductibles.

Pour le moment, aucune séquence centromérique candidate n'a pu être mise en évidence. Cependant, la recherche de ces séquences ne peut se faire que sur des séquences connues et assemblées (génome MAC, IES et contig mic). L'assemblage des régions mic-spécifiques étant encore très fragmenté, l'identification de séquences enrichies dans le génome micronucléaire est très limitée. A l'avenir, l'augmentation des rendements de production d'ADN micronucléaire et l'amélioration des techniques d'assemblage devrait favoriser l'analyse des données de ChIP et permettra peut-être de mettre en évidence des séquences d'ADN associées à CenH3.

Table 1. Quantité d'ADN récupérée dans les échantillons issus des expériences de ChIP 1 et 2 (dosage Qbit assay Invitrogen).

	Echantillon	Quantité d'ADN (ng)
Exp 1	ChIP sérum pré-immun	12,84
	ChIP CenH3 (4,5 ug)	20,88
	ChIP CenH3 (2ug)	38,34
Exp 2	ChIP IgG	22,2
	ChIP CenH3 (3 ug)	22,6
	ChIP H3K9me3 (5 ug)	15, 5

Table 2. Nombre de lectures alignées sur le génome MAC dans les échantillons Input, CenH3 et contrôle des expériences 1 et 2.

	Echantillon	Nombre de lectures (M=Millions)	Nombre de lectures alignées sur le génome MAC	Pourcentage de lectures alignées sur le génome MAC (%)
Exp 1	Input	8,6 M	7,7 M	88
	CenH3	6,0 M	5,2 M	86
	Pré-immun	8,3 M	7,3 M	88
Exp 2	Input	17,8 M	17 M	95
	CenH3	13,6 M	13,1 M	96
	IgG	17 M	16,3 M	96

CONTRIBUTIONS ADDITIONNELLES PENDANT LA THESE

1. Au cours de ma thèse, j'ai participé à l'étude qui a mis en évidence les bases moléculaires de la détermination et de l'hérédité du type sexuel chez *Paramecium tetraurelia* (ANNEXE 1, Singh et al., *Nature*, accepté). Dans cet article, j'ai contribué à l'analyse des gènes requis pour l'excision du promoteur mtA en mettant en évidence par PCR que les gènes NOWA1 et DCL2 et 3 étaient requis pour l'excision de mtA et l'hérédité cytoplasmique du phénotype O (Figure 4).

2. J'ai également participé à l'écriture d'une revue sur les mécanismes d'élimination d'ADN guidés par des ARN non codants chez *Paramecium tetraurelia* et *Tetrahymena thermophila* : Coyne RS*, **Lhuillier-Akakpo M***, Duharcourt S. RNA-guided DNA rearrangements in ciliates : is the best genome defense a good offense ? Biol Cell. 2012 Jun;104(6):309-25. (ANNEXE 2)

3. Enfin, j'ai participé à une cinétique de développement qui a permis d'obtenir des ARN totaux de cellules sauvages de *Paramecium* à différents stades du cycle de vie des cellules. Le séquençage des ARNm va permettre d'obtenir les profils d'expression de l'ensemble gènes de *Paramecium* pendant le cycle végétatif et le cycle de reproduction. Cela permettra de confirmer et de compléter les données de microarray déjà disponibles. L'analyse des ARNm permettra également de parfaire l'annotation du génome macronucléaire (Arnaiz et al. *en préparation*).

DISCUSSION

DISCUSSION

Chez *Paramecium tetraurelia*, la différenciation macronucléaire est associée à des réarrangements importants du génome qui conduisent à l'élimination de multiples séquences d'ADN, uniques, répétées, diverses en taille et en composition. Pour mieux comprendre comment ces séquences sont ciblées par l'endonucléase PiggyMac, nous avons cherché à déterminer si la chromatine avait un rôle à jouer dans le processus d'élimination d'ADN et notamment si des modifications d'histones étaient impliquées.

I. Ezl1p, une protéine Polycomb actrice de l'élimination programmée d'ADN chez *Paramecium*

1. Méthylation des histones et élimination imprécise des régions mic-spécifiques

Nous avons observé par des expériences d'immunofluorescence que la triméthylation de H3K27 et de H3K9 apparaissait transitoirement dans les macronoyaux en développement, de manière concomitante avec les événements d'élimination d'ADN. La déposition de ces marques d'histone dépend d'une lysine méthyltransférase putative apparentée aux protéines de la famille Enhancer of zeste, nommée Ezl1p. Si la lysine 27 de l'histone H3 est un substrat attendu pour une lysine méthyltransférase de type Enhancer of zeste, il n'y a pas d'évidence chez les eucaryotes supérieurs que ces protéines puissent méthyler la lysine 9 *in vivo*. Il est possible que l'inactivation de EZL1 ait un effet indirect sur la triméthylation de H3K9 en empêchant, par exemple, le recrutement d'une autre lysine méthyltransférase. Cependant, des données chez la *Drosophile* montrent que E(z) possède une double spécificité de substrat *in vitro* (K27 et K9), sur des queues d'histones H3, des octamères reconstitués et des polynucléosomes (Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002). Chez le cilié *Tetrahymena thermophila*, les mutants EZL1 présentent également un défaut de triméthylation sur H3K9 pendant le développement (Liu et al., 2007). De plus, nous n'avons pu trouver aucune lysine méthyltransférase spécifique de K9 dans le génome de *Paramecium*, de *Tetrahymena* ou d'*Oxytricha* (un autre cilié), suggérant que chez les

ciliés, les protéines Enhancer of zeste pourraient avoir la capacité de catalyser la triméthylation de K27 et de K9.

Par une analyse à l'échelle du génome entier, nous avons montré que EZL1 était nécessaire à l'élimination imprécise de l'ensemble des séquences mic-spécifiques, représentant environ 25 Mb du génome micronucléaire. EZL1 est également requis pour l'excision précise d'un sous-ensemble d'IES (cet aspect sera discuté plus tard). L'endonucléase PiggyMac n'est pas requise pour la déposition de H3K27me3 et H3K9me3 et agirait donc en aval de l'activité de EZL1. Bien que nous n'ayons pas de données de ChIP permettant de localiser ces modifications d'histone sur le génome, ces données suggèrent fortement que Ezl1p est requise pour la déposition de H3K9me3 et H3K27me3 au niveau des séquences à éliminer et que ces marques sont nécessaires, directement ou non, pour le recrutement de l'endonucléase PiggyMac.

Nous avons également confirmé à l'échelle du génome entier que l'élimination imprécise des séquences mic-spécifiques dépendait des protéines Dicer-like 2 et 3 (Dcl2p et Dcl3p), responsables de la biogenèse des scanARN en méiose, ce qui avait déjà été observé par Lepère *et al.* en 2008 sur une région de fragmentation des chromosomes. Parallèlement à cela, nous avons pu montrer que l'inactivation de DCL2 et DCL3 conduisait à une forte diminution de la quantité de H3K9me3 et H3K27me3 dans les macronoyaux en développement suggérant que les protéines Dcl2p, Dcl3p et Ezl1p interviennent dans une même voie et que les scanARN agissent en amont de l'activité de Ezl1p.

De manière similaire au recrutement de la lysine méthyltransférase *Clr4* par les siARN au niveau des régions péricentromériques chez *S. Pombe* et en accord avec le modèle de scanning proposé chez *Paramecium* et *Tetrahymena*, nous suggérons que les scanARN, en complexes avec les protéines Ptiwi01p et Ptiwi09p et portant des séquences mic-spécifiques, pourraient permettre de recruter Ezl1p au niveau des séquences à éliminer dans le macronoyau en développement. Ezl1p catalyserait alors la triméthylation de H3K27 et de H3K9 sur des nucléosomes associés aux séquences mic-spécifiques et permettrait l'attraction de PiggyMac dans ces régions. Il est possible qu'un mécanisme permette de limiter le recrutement de PiggyMac aux extrémités de la région à éliminer pour que celle-ci soit excisée en un seul fragment. Des cassures pourraient également être introduites en plusieurs points de la séquence et la région éliminée sous forme de plusieurs fragments. Aucune précision de coupure n'étant requise, les limites de la propagation des marques d'histones peuvent être variables (Figure 23a). Le même type de modèle a été proposé chez

Tetrahymena où il a également été observé que l'absence de scanARN affectait la triméthylation de H3K27 et de H3K9 médiée par Ezl1p (Liu et al., 2004, 2007). Chez les eucaryotes supérieurs, il n'existe aucune évidence que des voies de l'ARN interférence soient requises pour le ciblage des protéines Polycomb. Cependant, à l'image de la potentielle dualité de substrat de Ezl1p (K9 et K27), les protéines Polycomb des ciliés pourraient être recrutées à leur séquences cibles par des mécanismes similaires aux protéines Su(var)3-9. Il a notamment été proposé que les deux voies impliquées dans la répression de la chromatine, la voie dépendante de H3K9me2/me3 et HP1 et la voie dépendante des protéines Polycomb, qui présentent de fortes similarités entre elles, aient une origine commune. Ces deux voies auraient pu rester unifiées chez les organismes ciliés tandis qu'elles se seraient séparées au cours des étapes de l'évolution pour agir dans des processus différents chez les eucaryotes supérieurs (Beisel and Paro, 2011; Hekimoglu and Ringrose, 2009).

Suite à l'action de Ezl1p, comment les modifications d'histone peuvent-elles permettre le recrutement de l'endonucléase ? PiggyMac contient un domaine PHD à son extrémité N-terminale, potentiellement capable de fixer des modifications d'histones. Bien que la plupart des protéines à domaine PHD décrites reconnaisse des formes méthylées de H3K4, comme la recombinase RAG2 impliquée dans la recombinaison V(D)J chez les mammifères, (Matthews et al., 2007), il existe quelques exemples de domaines PHD interagissant avec la méthylation de H3K9. Par exemple, le facteur ICBP90 impliqué dans la formation ou la maintenance de l'hétérochromatine chez l'Homme fixe H3K9me3 au niveau des régions péri-centromériques par l'intermédiaire de son motif PHD (Karagianni et al., 2008). La protéine à chromodomaine CDH4 appartenant au complexe NuRD (Nucleosome remodeling and histone deacetylase) possède également deux domaines PHD qui lient H3K9me3 avec une forte affinité *in vitro* (Mansfield et al., 2011; Musselman et al., 2009). Par ailleurs, il semble que la triméthylation de H3K9 et de H3K27 renforce l'interaction avec H3 du domaine PHD de Tpb2p, l'homologue de PiggyMac chez *Tetrahymena* (Vogt and Mochizuki, 2013). Chez *Paramecium*, PiggyMac pourrait donc directement interagir avec les modifications d'histones déposées au niveau des séquences à éliminer (Figure 23b). L'endonucléase pourrait également être recrutée par des protéines agissant en aval de la méthylation et requises pour la formation d'une structure chromatiniennne particulière. PiggyMac possède notamment un domaine coiled-coil qui lui permettrait d'interagir avec d'autres protéines. Les méthylations de H3K27 et H3K9 sont connues pour servir de plateforme à la fixation de protéines à chromodomaines permettant, directement ou non, la condensation de la chromatine et l'installation d'un état de répression transcriptionnel.

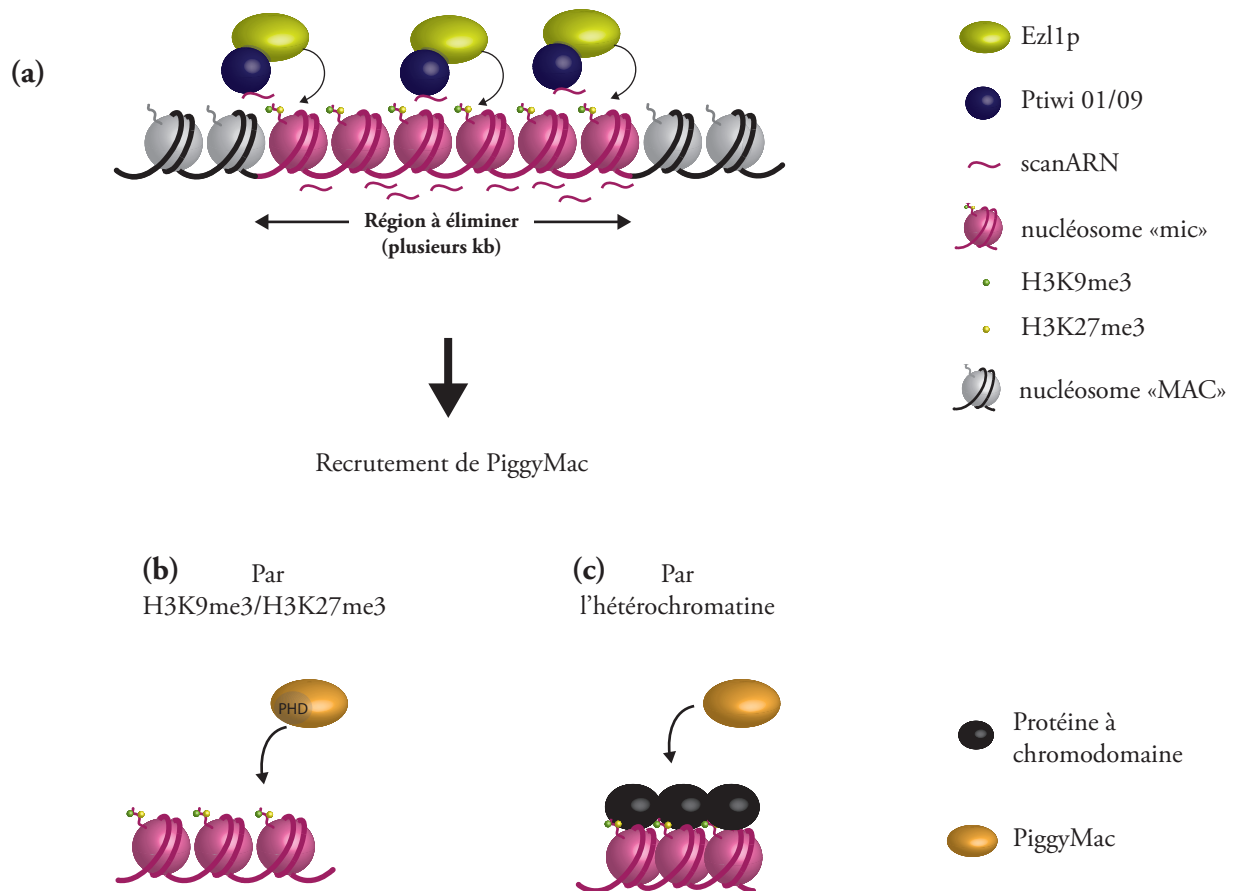


Figure 23. Modèle du mécanisme d'élimination imprécis des séquences micronucléaires. (a) Les scanARN, complexés aux protéines Ptiwi 01 et 09 et portant des séquences homologues à la région à éliminer, permettraient le recrutement de la lysine méthyltransférase Ezh1p. Celle-ci catalyserait la triméthylation de H3K27 et permettrait, directement ou non, la triméthylation de H3K9 sur les nucléosomes de la séquence micronucléaire. (b) L'endonuclease PiggyMac pourrait être recrutée directement au niveau de la région à éliminer par une interaction entre son domaine PHD et les histones modifiés. (c) L'endonuclease pourrait être recrutée par une structure hétérochromatinienne, formée grâce à l'interaction entre les histones modifiés et des protéines à chromodomaine.

H3K9me2 ou me3 interagit avec des protéines de la famille HP1 tandis que H3K27me3 se lie à des protéines à chromodomaine du complexe PRC1. Chez *Tetrahymena*, les protéines à chromodomaine Pdd1p et Pdd3p se lient à H3K9me3 au niveau des séquences à éliminer, Pdd1p reconnaît aussi H3K27me3. A l'instar de HP1, ces facteurs permettraient la formation d'une structure hétérochromatinienne à la périphérie des macronoyaux en développement (Liu et al., 2007; Madireddi et al., 1996; Taverna et al., 2002). Le ciblage artificiel de Pdd1p dans une région macronucléaire suffit à son élimination indiquant que la formation d'hétérochromatine apporte le signal ultime qui déclenche l'élimination (Taverna et al., 2002). Il est donc probable que chez *Paramecium*, les méthylations de H3K27 et de H3K9 contribuent au ciblage de protéines régulatrices de la formation d'hétérochromatine et que celle-ci soit nécessaire au recrutement de PiggyMac (Figure 23c). Il serait intéressant de caractériser les nombreuses protéines à chromodomaine qui existent dans le génome de *Paramecium* et les facteurs interagissant avec Ezl1p pour valider cette hypothèse.

2. Méthylation des histones et excision des IES

Nous avons montré que la lysine méthyltransférase Ezl1p était également requise pour l'excision précise de plus de 30 000 IES. Ces séquences ont des caractéristiques bien définies dans le génome de la paramécie : elles sont uniques, plutôt courtes (26 pb-5kb) et doivent être éliminées de manière très précise, au nucléotide près. De manière intéressante, toutes les catégories de taille d'IES sont affectées par l'inactivation de EZL1, à différents degrés (discuté plus bas), mais il existe des IES aussi courtes que 26 pb qui dépendent de EZL1 pour être excisées. Si le ciblage par Ezl1p d'IES de plusieurs kb peut facilement être comparé à celui des régions imprécisément éliminées, comment la méthylation des histones peut-elle réguler l'excision de séquences très courtes, plus courtes que la taille de l'ADN enroulé autour d'un nucléosome ? Il est possible d'imaginer une action beaucoup plus locale de la lysine méthyltransférase qui viendrait méthyler le ou les nucléosomes à proximité de l'IES à éliminer (Figure 24A). Dans cette situation, il devient difficile de penser que cette méthylation, très ciblée, pourrait conduire à la formation d'une structure hétérochromatinienne comme cela a été envisagé pour les séquences les plus longues. Le recrutement de l'endonucléase devrait donc se faire indépendamment de la formation d'une structure chromatinienne particulière. PiggyMac pourrait être recrutée directement par les

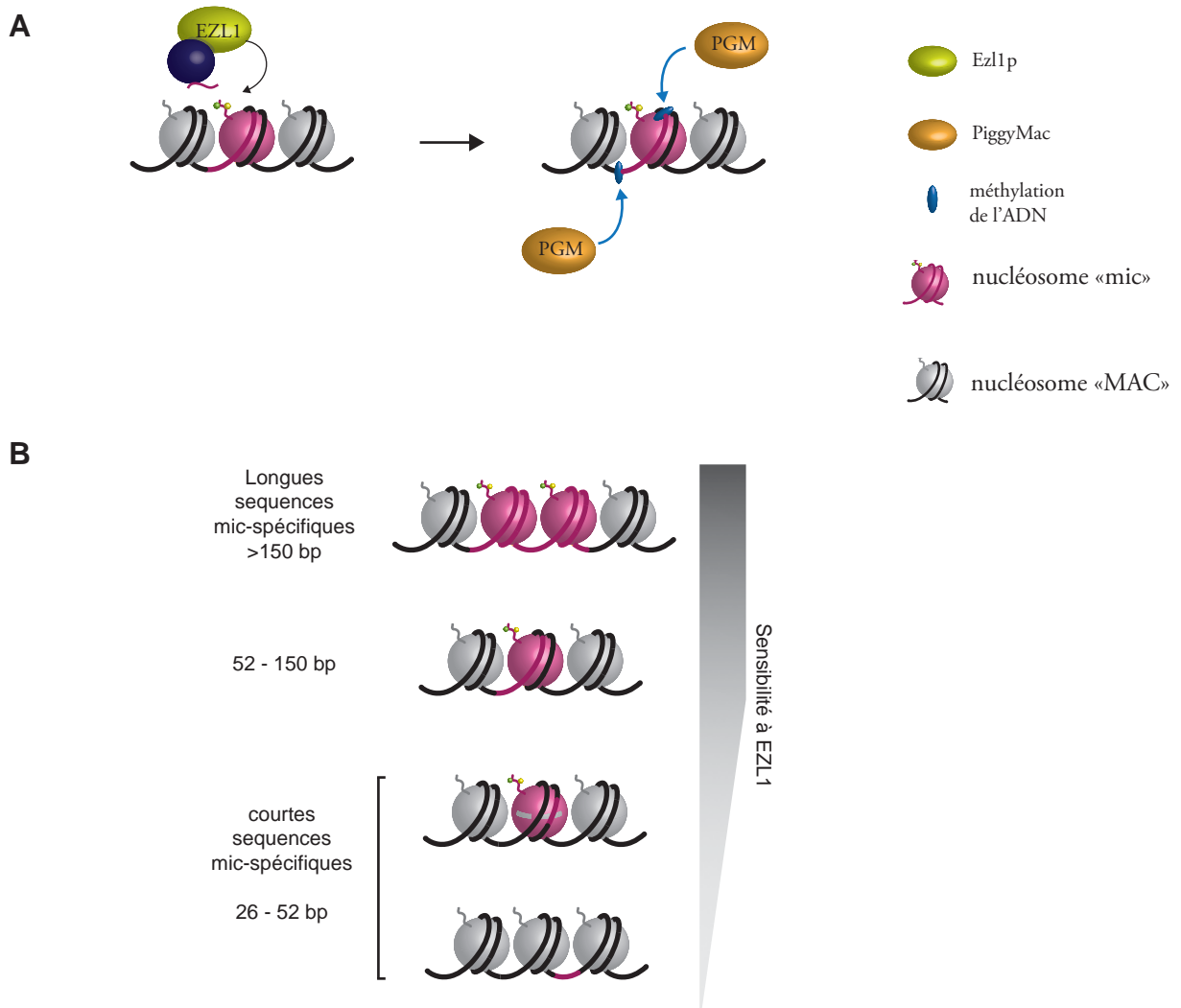


Figure 24. A. Au niveau des IES, Egl1p pourrait agir très localement et méthyler le nucléosome associé à la séquence à éliminer. Des modifications de nucléotides pourraient faciliter la précision de la coupure par PiggyMac. **B.** Modèle représentant les différents contextes nucléosomiques dans lesquels peuvent se trouver les IES. Jusqu'à 52 pb, les IES sont fortement sensibles à Egl1 car elles sont complètement ou partiellement incluses dans un nucléosome et leur reconnaissance dépend de la méthylation des histones. Entre 26 et 52 pb, les IES peuvent être incluses ou non dans un nucléosome expliquant la variabilité de leur dépendance à Egl1.

marques d'histones par l'intermédiaire de son domaine PHD comme décrit précédemment ou indirectement par le biais d'autres facteurs agissant en aval des modifications d'histone.

Même si elle est nécessaire pour le ciblage de PiggyMac au niveau des IES, la méthylation des histones ne peut assurément pas suffire pour conduire la précision de coupure qui est exigée pour l'élimination des IES. En effet, la précision du ciblage par un nucléosome a une résolution limitée à 147 pb, la longueur de l'ADN enroulé autour d'un nucléosome. D'autres mécanismes doivent donc être impliqués dans « l'annotation » des extrémités des IES pour augmenter la précision du ciblage de PiggyMac. Le consensus dégénéré retrouvé aux extrémités des IES (5'-TAYAGYNR-3') pourrait médier la précision de coupure si l'on envisage que PiggyMac possède une préférence de séquence pour catalyser les cassures de l'ADN. Les modifications de nucléotides sont également de bonnes candidates pour diriger la précision de l'excision. La modification la plus répandue dans les génomes eucaryotes est la méthylation des cytosines dans un contexte CpG qui est communément associée à la répression transcriptionnelle des gènes (Bird, 2002). Il existe notamment des exemples d'interaction entre la voie Polycomb et la méthylation de l'ADN. Il a été montré dans des cellules humaines que les ADN méthyltransférases DNMT1, DNMT3A et DNMT3B interagissaient avec EZH2 dans le complexe PRC2 pour être recrutées aux sites de répression contrôlés par les Polycomb (Viré et al., 2006). Chez *Paramecium*, il n'y a à ce jour aucune évidence de la présence de cytosine méthylée dans le génome. En revanche les séquences d'ADN macronucléaire contiennent des adénines N6-méthylées, une modification que l'on retrouve dans les génomes des bactéries, des archées et des champignons (Cummings et al., 1974). L'apparition de méthylation *de novo* des adénines aux sites GATC dans le macronoyau en développement des cellules de *Tetrahymena* au moment où se produisent les réarrangements du génome a conduit à penser que cette marque pourrait contribuer au ciblage des séquences à éliminer (Harrison and Karrer, 1985). Bien qu'il n'y ait aucune évidence de cette hypothèse chez *Tetrahymena* ou chez *Paramecium*, il est possible d'imaginer un modèle dans lequel Ezl1p recruterait des ADN méthyltransférases au niveau des IES. Celles-ci méthyleraient, par exemple, l'adénine à chaque extrémité des IES et faciliteraient la précision de la coupure (Figure 24A).

II. Des voies partiellement chevauchantes impliquées dans l'élimination d'ADN chez *Paramecium* : complexification du modèle de scanning

1. EZL1 affecte un sous-ensemble des IES sensibles à PGM

L'analyse du génome de cellules inactivées pour EZL1 a révélé qu'un sous-ensemble d'IES n'était pas retenu, de manière significative et reproductible, par rapport au génome de cellules contrôles. Le fait que certaines IES soient normalement excisées dans les conditions d'inactivation de EZL1 pourrait s'expliquer par une efficacité partielle de l'inactivation par ARN interférence de EZL1. Même si nous avons montré que l'inactivation de EZL1 conduisait à une réduction importante des ARNm correspondant à ce gène et générerait une réduction drastique des niveaux de triméthylation de H3K27 et H3K9, nous ne pouvons pas exclure qu'il reste une fraction de protéine Ezl1 active suite à l'inactivation de ce gène par ARN interférence. Cependant, la variabilité des scores de rétention entre les différentes IES pour une dose donnée et réduite de Ezl1p montre clairement une différence de sensibilité des IES à Ezl1p pour être excisées. Sur la base de paramètres statistiques, nous avons pu distinguer deux sous-ensembles d'IES : les IES significativement retenues suite à l'inactivation de EZL1 et les IES non significativement retenues. Comme évoqué précédemment pour la précision de coupure des IES, ces données confirment qu'il doit exister un mécanisme de ciblage des IES en marge de la méthylation des histones. Très peu de critères au niveau génomique nous ont permis de distinguer les IES sensibles à EZL1 des IES insensibles, suggérant que les informations discriminantes entre ces deux classes d'IES se trouvent à une autre échelle, structurale ou temporelle.

a. Dépendance à EZL1 et organisation de la chromatine

Le principal critère permettant de distinguer les deux catégories d'IES est leur taille, les IES les plus longues étant les plus sensibles à EZL1 avec un seuil à 52 pb au dessus duquel quasiment toutes les IES (90%) sont sensibles à la lysine méthyltransférase. Cette caractéristique structurale pourrait être directement liée à l'organisation de la chromatine et refléter la position des IES par rapport aux nucléosomes. En effet, plus une IES est longue, plus elle a de chance d'être comprise dans un fragment d'ADN qui est inclus dans un nucléosome et donc plus elle dépendrait de Ezl1p pour être reconnue et ciblée par PiggyMac. On peut même imaginer qu'au dessus de 52

pb, une IES sera obligatoirement incluse, partiellement ou entièrement, dans un nucléosome et aura donc un niveau de dépendance à EZL1 équivalent à celui d'une IES de plusieurs centaines de pb pour être excisée. Cette hypothèse sous-entend que la taille des régions linker – séquences d'ADN libre qui séparent les nucléosomes – dans le macronoyau en développement est inférieure à 52 pb. Aucune information sur la taille de ces régions ou la position des nucléosomes dans les génomes des ciliés n'est actuellement connue mais chez les eucaryotes pour lesquels cela a été décrit la taille des régions linker peut varier entre 20 et 90 pb (Szerlong and Hansen, 2011). De manière frappante, entre 26 et 52 pb, les niveaux de rétention des IES sont très variables à l'intérieur d'une même classe de taille. Cela pourrait s'expliquer par le fait que ces IES peuvent se trouver dans des contextes très différents en étant soit entièrement incluses dans un nucléosome, soit partiellement incluses, soit complètement indépendante d'un nucléosome. Dans chacune de ces situations, la méthylation des histones devrait coopérer, de manière plus ou moins importante, avec d'autres mécanismes pour permettre le ciblage de PiggyMac. Les différences de scores de rétention en absence de EZL1 pourraient donc refléter la contribution plus ou moins importante apportée par la méthylation des histones pour la reconnaissance des IES en fonction du contexte nucléosomique dans lequel se trouvent ces IES (Figure 24B). Cette situation rappelle ce qui a été décrit chez *S. cerevisiae* pour le recrutement du facteur de remodelage de la chromatine SWR1 au niveau des régions promotrices. *In vitro*, SWR1 fixe préférentiellement les longues régions d'ADN dépourvues de nucléosome mimant la chromatine des promoteurs tandis que l'acétylation des histones dans cette chromatine favorise la fixation du facteur. Le signal porté par l'ADN est « dominant » par rapport à celui de l'histone modifiée mais la coopération hiérarchique entre les deux signaux améliore l'efficacité du recrutement de SWR1 (Ranjan et al., 2013). Chez *Paramecium*, la balance entre méthylation des histones et un autre mécanisme pourrait accroître l'efficacité du recrutement de PiggyMac.

A l'avenir, l'analyse par CHIP de la localisation des marques H3K9me3 et H3K27me3 dans le génome du macronoyau en développement, conjointement avec la mise au point de techniques permettant d'étudier la position des nucléosomes dans le génome macronucléaire en développement, devraient nous permettre de tester ce modèle.

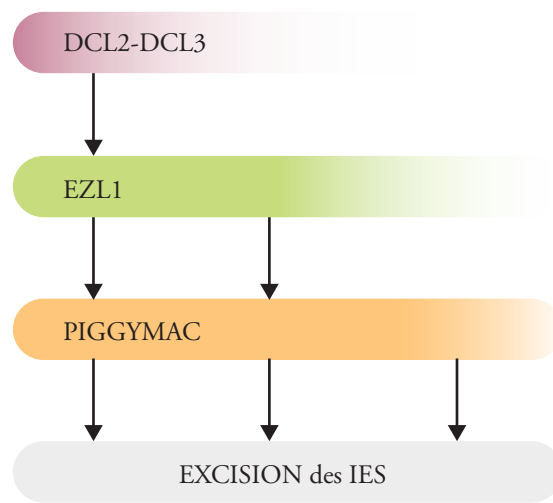


Figure 25. Les différentes voies impliquées dans l'élimination des IES chez *Parametium*. Une voie, similaire à l'élimination imprécise des longues séquences micronucléaires, implique les protéines productrices des scanARN Dcl2p et Dcl3p et la lysine méthyltransférase Ezl1p. Une deuxième voie implique Ezl1p indépendamment des scanARN et une troisième voie ne requière ni les scanARN ni Ezl1p mais des mécanismes encore inconnus. Toutes ces voies convergent vers le recrutement de PiggyMac pour l'excision des IES.

b. Dépendance à EZL1 et programme temporel d'élimination

Si elles n'étaient pas liées à l'organisation des nucléosomes, les différences de sensibilités à EZL1 pourraient s'expliquer par une différence d'action de la protéine Ezl1 dans le temps. Les IES les plus sensibles à EZL1 seraient excisées dans une même fenêtre de temps pendant laquelle Ezl1p aurait un maximum d'activité. On ne connaît pas le timing d'élimination des IES à l'échelle du génome entier mais, de manière cohérente avec les effets de EZL1 sur la rétention des IES, la taille des séquences à éliminer pourrait être corrélée au moment de leur excision. Il existe notamment des exemples d'IES imbriquées où une petite IES est excisée avant la grande IES dans laquelle elle est incluse (Duharcourt et al., 1998; Mayer et al., 1998). Il existe notamment une petite IES imbriquée dans l'IES A6649 et nous avons pu observer que la grande IES était dépendante de EZL1 tandis que la petite IES était insensible à EZL1 pour être excisée (voir article 1, Figure S9). Il pourrait donc exister un programme d'élimination régulé dans le temps qui ferait intervenir séquentiellement différents facteurs. Les IES insensibles à EZL1 seraient des IES « précoces », excisées par un mécanisme indépendant de EZL1 avant que la lysine méthyltransférase ne soit active, tandis que les IES sensibles à EZL1 correspondraient à des IES plus « tardives ».

2. DCL2 et DCL3 affectent un sous-ensemble des IES sensibles à EZL1

L'analyse de la rétention des IES dans les conditions d'inactivation de DCL2 et DCL3 montre que seulement 7% de ces séquences sont significativement retenues. Les IES significativement affectées par la déplétion de DCL2 et DCL3 sont toutes incluses dans le sous-ensemble d'IES sensibles à EZL1. Elles font partie des IES les plus longues et les plus sensibles à EZL1, confirmant que DCL2, DCL3 et EZL1 doivent faire partie d'une même voie. Cependant la différence de rétention importante entre les deux conditions suggère qu'il existe une voie d'élimination dépendante de EZL1 et indépendante des scanARN. Cette différence entre les effets de EZL1 et de DCL2 et DCL3 est-elle biologique ou s'explique-t-elle par une faible efficacité de l'inactivation de DCL2 et DCL3 ? Il est très difficile de comparer le niveau d'inactivation de DCL2 et DCL3 par rapport à celui de EZL1. Cependant, l'effet fort de l'inactivation de DCL2 et DCL3 sur la méthylation des histones et sur la rétention des séquences éliminées imprécisément et le fait que le même résultat sur les IES ait été obtenu par un autre

groupe (Sandoval et al., 2014) sont autant d'arguments qui soutiennent la bonne efficacité de l'inactivation de DCL2 et DCL3 par ARN interférence. Les IES semblent donc très peu sensibles à l'action des scanARN suggérant que Ezl1p serait ciblée par un mécanisme indépendant des scanARN.

Si elle n'est pas recrutée par les scanARN, d'autres petits ARN non codants pourraient être impliqués. Nous avons notamment envisagé que des petits ARN produits par l'activité slicer des protéines Ptiwi01 et 09, indépendamment des protéines Dicer-like, pourraient jouer un rôle dans le recrutement de Ezl1p. Nos données préliminaires concernant le séquençage du génome de cellules inactivées pour PTIW101 et PTIW109 ne vont pas tellement dans ce sens mais nos résultats sont difficiles à interpréter sans duplicat biologique. Il est donc envisageable que des petits ARN produits par les protéines Ptiwi permettent le recrutement de Ezl1p indépendamment des scanARN mais il pourrait aussi exister des voies indépendantes de l'ARN interférence pour recruter Ezl1p au niveau des IES de *Paramecium*.

En conclusion, l'ensemble de nos données met en évidence des différences de contribution entre les différents facteurs impliqués dans les réarrangements du génome chez *Paramecium*. Il existerait au moins trois voies, partiellement chevauchantes, qui participeraient à l'élimination précise et imprécise des séquences micronucléaires. L'implication de chacune de ces voies pourrait être influencée par le contexte génomique, l'organisation des nucléosomes dans la chromatine ou encore du timing d'excision (Figure 25). Il est difficile de savoir si, au niveau cellulaire, ces voies sont représentées par des complexes protéiques différents ou s'il existe un complexe d'excision majeur dont chacun des facteurs joue un rôle plus ou moins importants en fonction des séquences à éliminer. A l'avenir, la combinaison d'approches génétiques et biochimiques permettra peut-être l'identification des partenaires de Ezl1p et de PiggyMac pour déterminer quels autres facteurs peuvent être impliqués dans le recrutement de ces protéines et quels sont les différents complexes d'excision. Chez les eucaryotes, différentes voies participent également à la formation d'hétérochromatine. Chez *S. pombe*, l'hétérochromatine constitutive peut être établie par la voie de l'ARN interférence au niveau des régions péri-centromériques ou par des facteurs de transcription séquence-spécifiques au niveau des télomères. Ces deux voies convergent vers le recrutement de la lysine méthyltransférase Clr4 (Su(var)3-9) et du chromodomaine Swi6 (HP1) pour la formation d'hétérochromatine. De manière similaire, chez différents organismes, la

répression par les protéines polycomb peut être mise en place par des séquences d'ADN spécifiques, des facteurs de transcription ou encore des ARN non codants (Beisel and Paro, 2011). Comment et pourquoi un si grand nombre de mécanismes permet d'accomplir une même fonction reste à élucider. Le cilié *Paramecium tetraurelia* représente donc un excellent modèle pour étudier les voies impliquant les méthylations de K9 et de K27 et pour mieux comprendre en quoi le contexte nucléaire peut influencer le choix d'une voie ou d'une combinaison de voies pour un but commun, le recrutement d'une endonucléase.

III. Rôles alternatifs de EZL1 et des scanARN

Et si Ezl1p et les scanARN n'étaient pas requis pour le ciblage de PiggyMac mais seulement pour l'action de l'endonucléase? Nous avons montré que EZL1 était requis en amont de l'activité de PiggyMac puisque l'inactivation de EZL1, en plus de conduire à la rétention des formes chromosomiques de certaines IES, conduit à l'absence de produit d'excision d'IES (article 1 Figure S6 : pas de cercle d'IES pour l'IES A4578 dans les conditions d'inactivation de EZL1). Cependant, nous n'avons pas montré que Ezl1p était requise en amont du recrutement de PiggyMac. Pour cela, il faudrait montrer par des expériences de ChIP que PiggyMac est délocalisée en absence de Ezl1p. Si elle n'était pas nécessaire au recrutement de PiggyMac, Ezl1p, par l'intermédiaire de la méthylation de H3K27me3 et H3K9me3 pourrait être impliquée dans le recrutement d'autres facteurs qui facilitent l'excision de l'ADN par PiggyMac. En effet, la présence de nucléosome pourrait non seulement être une contrainte pour la reconnaissance des IES mais aussi être une entrave pour la coupure de l'ADN par l'endonucléase. Il est tout à fait probable que des facteurs de remodelage de la chromatine soient impliqués dans l'élimination d'ADN. La méthylation des histones pourrait par exemple permettre le recrutement de protéines HMG (High Mobility Group), ces facteurs « fluidifient » la chromatine et facilitent l'accès à l'ADN entouré autour des nucléosomes (Bianchi and Agresti, 2005). Ces protéines HMG seraient requises principalement pour les IES associées à des nucléosomes pour faciliter l'accès de PiggyMac à l'ADN et seraient donc requises pour l'excision des IES sensibles à EZL1. Ce modèle pourrait expliquer les différences de timing d'excision qui existent entre les différentes séquences à éliminer, les séquences les plus courtes étant moins contraintes par la structure de la chromatine et donc plus rapidement éliminées.

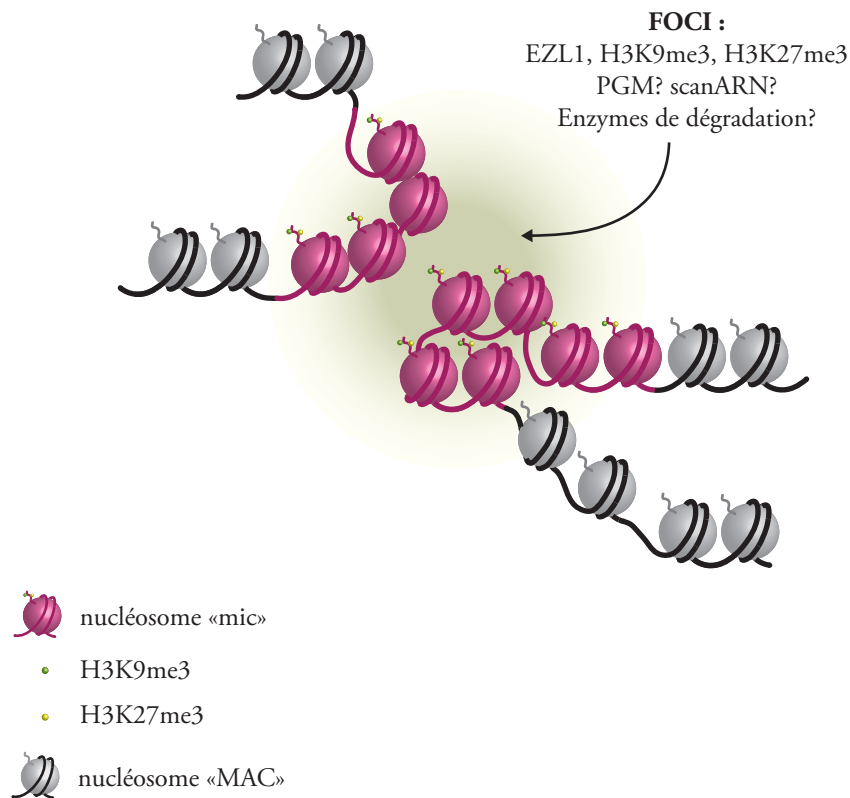


Figure 26. Modèle expliquant la formation de foci EZL1, H3K27me3 et H3K9me3 au cours du développement macronucléaire. Différentes séquences micronucléaires à éliminer seraient regroupées dans certaines régions du noyau formant des zones de concentration de Ezl1p, H3K27me3 et H3K27me3. PiggyMac pourrait participer à la formation de ces structures qui permettraient d'augmenter l'efficacité de l'élimination des régions mic-spécifiques. Ces foci pourraient également se former après l'excision pour faciliter la dégradation des séquences micronucléaires.

De même, le rôle des scanARN n'est peut-être pas en amont du recrutement de Ezl1p et de PiggyMac, même si ceux-ci sont nécessaires à l'activité de ces protéines. Il a été envisagé que la machinerie du RNAi pourrait être impliquée dans la formation de structures chromatinienne d'ordre supérieur en agissant comme une « glue » entre des loci hétérochromatinisés ou entre les loci ciblés par les protéines PcG (Grimaud et al., 2006; Hall et al., 2003).

De manière intéressante, nous avons observé que la triméthylation de H3K27 et de H3K9 présentait, avec Ezl1p, une localisation dynamique au cours du développement macronucléaire. D'abord diffus dans les macronoyaux en développement, ces facteurs se concentrent en foci au fil du développement et semblent former des structures nucléaires d'ordre supérieur similaires aux « polycomb bodies » décrits chez la Drosophile ou les cellules humaines (Pirrotta and Li, 2012), avant de disparaître. Les scanARN ou d'autres facteurs de la machinerie de l'ARN interférence pourraient être requis pour la mise en place de ces structures chromatinienne d'ordre supérieur. Cette hypothèse alternative au rôle des petits ARN non codants pourrait expliquer pourquoi, dans notre système, seules les séquences les plus longues (longues IES et régions imprécisément éliminées) susceptibles de former des structures hétérochromatiniennes, sont affectées par l'inactivation de DCL2 et DCL3. De manière cohérente avec cette hypothèse, l'inactivation de DCL2 et DCL3 n'affecte pas la localisation de Ezl1p dans les macronoyaux en développement mais altère la formation des foci Ezl1p. De plus, il a été observé que la localisation de la protéine Dcl5, nécessaire à la biogenèse des iesARN, était également ponctuée dans les macronoyaux en développement (Sandoval et al., 2014) suggérant que les voies de l'ARN interférence pourraient avoir un rôle dans la mise en place ou le maintien des structures chromatinienne d'ordre supérieur.

PiggyMac est requis pour la formation des foci suggérant que ces structures nucléaires se forment en aval du recrutement de l'endonucléase. Il serait intéressant de déterminer si l'activité catalytique de PiggyMac est requise pour la formation de ces structures ou si celles-ci se forment en amont l'excision de l'ADN. L'apparition des foci pourrait donc représenter la mise en place d'une organisation nucléaire particulière regroupant différentes régions mic-spécifiques. Elles pourraient améliorer l'efficacité du processus d'excision ou de dégradation de ces séquences (Figure 26).

Si les scanARN et Ezl1p étaient seulement nécessaires à l'action de coupure de PiggyMac et non pas à son ciblage, il faudrait envisager que PiggyMac puisse être recrutée par des mécanismes

différents, impliquant des facteurs qui n'ont pas encore été identifiés. Une hypothèse intéressante consiste à penser que les séquences à éliminer seraient prédéfinies dans les micronoyaux. En effet, au moment où les macronoyaux acquièrent leurs fonctions somatiques et subissent tous ces réarrangements génomiques, les nouveaux micronoyaux développent des fonctions germinales et certains mécanismes pourraient participer à l'annotation des séquences à éliminer dans le génome micronucléaire, pour faciliter l'élimination de ces séquences dans la macronoyau à la prochaine génération. L'étude du développement des micronoyaux étant limitée par les moyens expérimentaux actuels, cette hypothèse n'a pas encore été explorée.

Cependant, si les séquences à éliminer pouvaient être reconnues dans les micronoyaux, on peut se demander pourquoi, alors, ne pas éliminer directement les IES, transposons et autres séquences répétées du génome micronucléaire et se débarrasser définitivement de ces séquences « parasites » ? Chez les eucaryotes supérieurs, les régions non codantes du génome et les éléments transposables ont pendant longtemps été considérées comme de l'ADN « poubelle » (« junk » DNA), inutile au fonctionnement des cellules. Pourtant, ces régions d'ADN révèlent progressivement leurs rôles dans la régulation de l'expression des gènes, la diversité et l'adaptabilité des cellules. De même, si la paramécie garde une copie germinale de cet ADN qu'elle met à la « poubelle » dans le noyau somatique, c'est certainement parce que ces séquences d'ADN possèdent une fonction sous-jacente.

Dans l'article *Genome defence small RNAs exapted for epigenetic mating type inheritance* (Singh *et al.*, annexe 1), il est notamment montré que les séquences mic-spécifiques peuvent contribuer à la régulation de l'expression génique. En effet, l'expression du gène *mtA*, qui détermine le type sexuel des cellules de Paramécie, est régulée par le réarrangement alternatif de son promoteur qui s'avère être une IES. Dans les cellules de type sexuel O, le promoteur de *mtA* est excisé de manière dépendante aux scanARN et le gène ne s'exprime pas tandis que dans les cellules de type sexuel E, le promoteur est maintenu. L'IES est régulée épigénétiquement expliquant l'hérédité cytoplasmique du type sexuel chez *Paramecium tetraurelia*.

Les séquences mic-spécifiques pourraient également participer à la répression de l'activité transcriptionnelle du génome micronucléaire, même si peu de choses sont connues à ce sujet. En outre, l'ADN micronucléaire pourrait être une source de diversité génétique dans les populations cellulaires car la rétention programmée de certaines séquences pourrait offrir une possibilité d'adaptation en réponse à des changements environnementaux. Enfin, les séquences d'ADN mic-

spécifiques pourraient jouer un rôle dans une des seules fonctions importante connue des micronoyaux : la transmission fidèle du matériel génétique d'une génération à l'autre. Cette mission repose sur les capacités du micronoyau à exécuter des divisions mitotiques pendant la croissance végétative et des divisions méiotiques pendant le cycle sexuel pour maintenir l'intégrité d'un génome qui servira à produire le génome somatique. La ségrégation fidèle des chromosomes à chaque division cellulaire est assurée par les centromères, des régions chromosomiques qui interagissent avec les microtubules mitotiques et permettent la séparation précise des chromatides sœurs. Les séquences mic-spécifiques pourraient donc participer à la fonction centromérique chez *Paramecium*. Pour approfondir cette hypothèse, une partie de ma thèse a consisté à étudier la régulation de l'activité des centromères chez *Paramecium*.

IV. Régulation de l'activité centromérique chez *Paramecium tetraurelia*

1. Les centromères : des séquences mic-spécifiques ?

Pour comprendre les mécanismes qui régulent l'activité des centromères chez *Paramecium*, nous avons identifié le variant d'histone centromérique CenH3 de notre organisme modèle. Nous avons montré que la fonction centromérique était restreinte aux micronoyaux. Ces données confirment les observations cytologiques de Tucker *et al.*, qui avaient déjà mis l'accent dans les années 1980 sur l'absence de fuseau mitotique et l'absence de condensation des chromosomes dans le macronoyau en division. Les chromosomes macronucléaires, dépourvus de centromère actif, seraient donc ségrégés aléatoirement entre les deux cellules filles. La forte ploïdie du macronoyau pourrait lui permettre de garder un nombre suffisant de copie de chaque chromosome même si la division n'a pas été fidèle. Cependant, aucune donnée au niveau moléculaire ne permet à ce jour de comprendre les mécanismes qui régulent la division des chromosomes macronucléaires.

L'absence de fonction centromérique dans le macronoyau végétatif s'explique par un phénomène d'inactivation des centromères germinaux dans le macronoyau en développement qui se traduit par la disparition du variant d'histone CenH3. L'inactivation des centromères pourrait s'expliquer soit par un processus épigénétique qui conduirait à l'éviction de CenH3 des régions centromériques, soit par l'élimination physique des séquences centromériques du génome

macronucléaire. Dans cette dernière hypothèse, l'ADN centromérique est, comme CenH3, restreint au génome micronucléaire.

Les expériences de ChIP dirigées contre la protéine CenH3 n'ont pas encore permis d'identifier des séquences centromériques candidates. Cependant, l'étude de gènes impliqués dans le processus d'inactivation des centromères permet d'avancer des hypothèses quant au mécanisme d'inactivation des centromères et à la nature des séquences centromériques. En effet, l'inactivation de l'endonucléase PiggyMac et de la lysine méthyltransférase EZL1 conduit à la rétention de CenH3 dans les macronoyaux à la fin du développement macronucléaire suggérant que l'inactivation des centromères dépend de l'élimination physique des séquences centromériques. En revanche, l'inactivation des centromères ne dépend pas des gènes DCL2 et DCL3. Nous avons montré précédemment que des voies partiellement chevauchantes étaient impliquées dans l'élimination d'ADN pendant le développement macronucléaire et nous avons notamment mis en évidence une voie dépendante de PiggyMac et de EZL1 mais indépendante des scanARN pour l'excision des IES. Alors les séquences centromériques sont-elles des IES insensibles à DCL2 et DCL3 ? Si tel était le cas, les nucléosomes CenH3 seraient répartis dans l'ensemble du génome et devraient former des chromosomes holocentriques. Les séquences des centromères pourraient également correspondre à des IES inconnues présentes dans les régions imprécisément éliminées pour former cette fois des chromosomes monocentriques. Il semble toutefois que le moment où se produit l'excision des IES (10-15 heures après le début du cycle sexuel, (Bétermier et al., 2000)) soit plus précoce que celui de l'inactivation des centromères (30-40 heures après le début du cycle sexuel). Cette observation est compatible avec l'idée que les centromères soient des IES car nous avons pu observer que certains produits d'excision des IES étaient maintenus dans la cellule sous forme circulaire avant d'être activement dégradés. Les régions centromériques pourraient donc être excisées des chromosomes macronucléaires avant que la chromatine associée ne soit activement dégradée par un processus de dégradation mis en place à la fin du développement. Le moment de l'inactivation des centromères, que nous décrivons comme la disparition de CenH3, pourrait donc être postérieur au moment de l'élimination des séquences centromériques.

Même si les données d'immunofluorescence avec l'anticorps anti-CenH3 semblent nous indiquer que les centromères font partie des séquences retenues dans le génome de cellules inactivées pour EZL1 et éliminées du génome de cellules inactivées pour DCL2 et DCL3, il est difficile de comparer des données de séquençage avec des données d'immunofluorescence qui ne sont

absolument pas quantitatives. En effet, la présence ou l'absence d'un signal en immunofluorescence pourrait traduire une simple différence des niveaux de rétention des séquences centromériques dans les différents génomes. Le seuil de rétention en dessous duquel le signal CenH3 ne sera plus visible en immunofluorescence est inconnu. Ainsi, les séquences centromériques pourraient tout aussi bien correspondre à des séquences répétées dans des régions imprécisément éliminées qui sont retenues dans le génome de cellules inactivées pour DCL2 et DCL3 mais moins fortement que dans le génome de cellules inactivées pour EZL1 ou PGM.

2. Mécanismes de l'inactivation des centromères

La figure 27 présente un modèle du mécanisme d'inactivation des centromères chez *Paramecium*. Le processus dépend de EZL1 et impliquerait donc la triméthylation de H3K27 et de H3K9. Chez les eucaryotes, les nucléosomes contenant l'histone canonique H3 intercalés entre les nucléosomes contenant CenH3 sont décrits comme étant diméthylés sur la lysine K4 (Sullivan and Karpen, 2004). Le processus d'inactivation des centromères pourrait donc faire intervenir des déméthylases pour éliminer ces marques et changer le statut de la chromatine des centromères à inactiver. De manière intéressante, il a été montré chez *S. pombe* que les régions centrales des centromères inactivés épigénétiquement (sans délétions d'ADN centromérique) sont associés à des facteurs de l'hétérochromatine (H3K9me2 et Swi6) et hypoacétylés en H3K14 et H3K9 marquant un changement d'état de la chromatine vers un état réprimé (Sato et al., 2012). Même tendance chez le maïs où l'ADN de centromères inactivés est hyperméthylé par rapport à celui des centromères actifs (Koo et al., 2011). Chez *Paramecium*, la triméthylation de H3K27 et de H3K9 médiée par Ezl1p permettraient le recrutement de PiggyMac, comme au niveau des autres séquences éliminées sensibles à EZL1, et l'élimination subséquente de l'ADN centromérique. Les régions centromériques pourraient être excisées des chromosomes macronucléaires et dégradées plus tardivement.

Il a été proposé chez *Tetrahymena* que le variant d'histone centromérique, CNA1, pourrait lui-même participer au processus d'élimination de l'ADN car les cellules mutantes pour CNA1 présentent des défauts d'élimination d'IES (Cui and Gorovsky, 2006). Chez *Paramecium*, l'inactivation par ARN interférence de CenH3-1 et CenH3-2 n'a pas conduit à la rétention des quelques séquences micronucléaires testées (non montré).

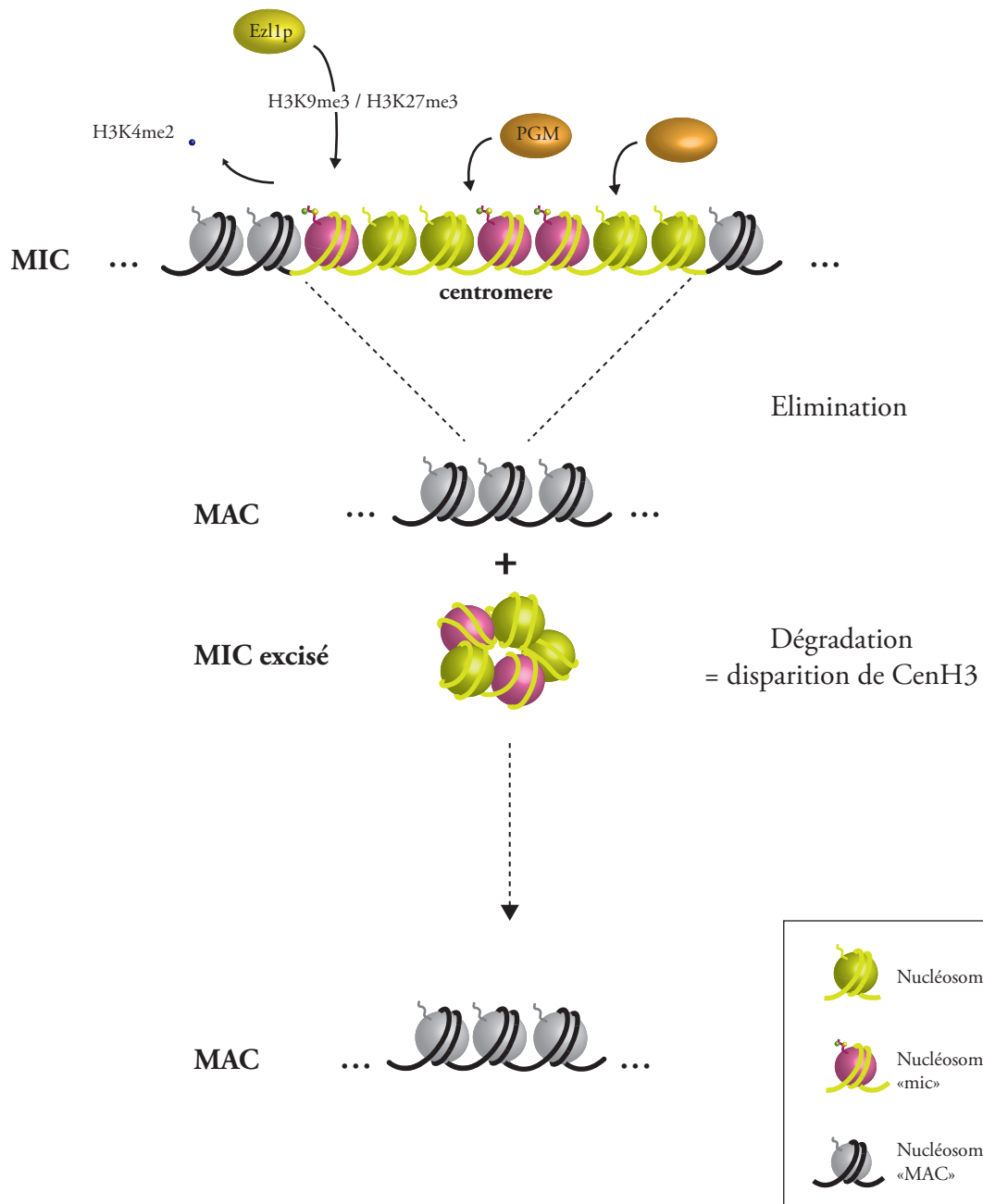


Figure 27. Modèle du mécanisme d'inactivation des centromères dans le macronoyau en développement chez *Paramecium*. Les séquences centromériques seraient ciblées par Ezl1p, indépendamment des scanARN et les nucléosomes canoniques, intercalés entre les nucléosomes CenH3, seraient modifiés par la triméthylation de H3K27 et H3K9. Des déméthylases pourraient venir éliminer des marques déjà présentes comme la diméthylation de H3K4. Les régions centromériques seraient ciblées par PiggyMac et physiquement éliminées du génome macronucléaire. Les régions centromériques excisées pourraient être maintenues un certain temps sous forme circulaire avant d'être dégradées.

Pour valider le modèle de l'inactivation par l'élimination physique des séquences, il faut identifier des séquences d'ADN centromérique de *Paramecium* et montrer leur absence du génome macronucléaire. En effet, le phénotype de rétention de CenH3 observé suite à l'inactivation de PiggyMac ou de EZL1 pourrait être causé par des effets indirects. L'inactivation de ces gènes pourrait bloquer la mise en place d'un processus requis pour l'inactivation épigénétique des centromères. Par exemple, la quantité de CenH3 dans les macronoyaux en développement pourrait être réduite par un contrôle de l'expression de CenH3 au niveau transcriptionnel, post-transcriptionnel ou traductionnel. Des modifications post-traductionnelles de CenH3 pourraient également déstabiliser son incorporation dans la chromatine ou encore, l'import de CenH3 dans les macronoyaux pourrait être diminué à un moment donné du développement. Il faut ajouter que ces deux modèles, inactivation épigénétique et inactivation par élimination des séquences centromériques ne sont pas mutuellement exclusifs. Chez *S. pombe* par exemple, les deux types de mécanismes peuvent se produire dans un même clone cellulaire (Sato et al., 2012). Il est donc possible d'imaginer que chez *Paramecium*, certaines régions centromériques sont physiquement éliminées, en conséquences de leur présence dans des séquences mic-spécifiques, tandis que d'autres sont épigénétiquement inactivées.

En conclusion, l'activité des centromères de *P. tetraurelia* est régulée au cours du développement. Un mécanisme d'inactivation des centromères germinaux, programmé, massif et reproductible est associé au développement macronucléaire et conduit à la disparition de la fonction centromérique dans les macronoyaux. Un certain nombre d'évidences tendent à montrer que les centromères de *Paramecium* sont inactivés dans le macronoyau par l'élimination physique des séquences centromériques. L'inactivation des centromères par délétion de l'ADN centromérique a été observée dans le contexte de chromosomes dicentriques chez *S. pombe* et chez l'Homme mais les mécanismes impliqués n'ont pas encore été élucidés (Sato et al., 2012; Stimpson et al., 2010). Il est tout a fait possible d'envisager que certaines voies impliquées chez *Paramecium* soient conservées chez les autres eucaryotes. Le cilié *P. tetraurelia* offre la possibilité d'étudier facilement le processus d'inactivation des centromères qui est plus rare à observer dans les autres systèmes. A l'avenir, l'identification de nouveaux gènes requis pour l'inactivation des centromères et l'identification des séquences centromériques de *P. tetraurelia* devrait permettre de percer certains mystères de la fonction centromérique chez les eucaryotes.

REFERENCES

- Agudo, M., Abad, J.P., Molina, I., Losada, A., Ripoll, P., and Villasante, A. (2000). A dicentric chromosome of *Drosophila melanogaster* showing alternate centromere inactivation. *Chromosoma* 109, 190–196.
- Allshire, R.C., and Karpen, G.H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat. Rev. Genet.* 9, 923–937.
- Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.-M., Denby Wilkes, C., Garnier, O., Labadie, K., Lauderdale, B.E., Le Mouël, A., et al. (2012). The *Paramecium* Germline Genome Provides a Niche for Intragenic Parasitic DNA: Evolutionary Dynamics of Internal Eliminated Sequences. *PLoS Genet* 8, e1002984.
- Aronica, L., Bednenko, J., Noto, T., DeSouza, L.V., Siu, K.W.M., Loidl, J., Pearlman, R.E., Gorovsky, M.A., and Mochizuki, K. (2008). Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in *Tetrahymena*. *Genes Dev.* 22, 2228–2241.
- Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., et al. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444, 171–178.
- Bailey, A.O., Panchenko, T., Sathyan, K.M., Petkowski, J.J., Pai, P.-J., Bai, D.L., Russell, D.H., Macara, I.G., Shabanowitz, J., Hunt, D.F., et al. (2013). Posttranslational modification of CENP-A influences the conformation of centromeric chromatin. *Proc. Natl. Acad. Sci.* 110, 11827–11832.
- Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in *Drosophila*. *Cell* 144, 214–226.
- Barnhart, M.C., Kuich, P.H.J.L., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., and Foltz, D.R. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J. Cell Biol.* 194, 229–243.
- Bassett, E.A., DeNizio, J., Barnhart-Dailey, M.C., Panchenko, T., Sekulic, N., Rogers, D.J., Foltz, D.R., and Black, B.E. (2012). HJURP Uses Distinct CENP-A Surfaces to Recognize and to Stabilize CENP-A/Histone H4 for Centromere Assembly. *Dev. Cell* 22, 749–762.
- Baudry, C., Malinsky, S., Restituto, M., Kapusta, A., Rosa, S., Meyer, E., and Bétermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev.* 23, 2478–2483.
- Beisel, C., and Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. *Nat. Rev. Genet.* 12, 123–135.
- Bergmann, J.H., Rodríguez, M.G., Martins, N.M.C., Kimura, H., Kelly, D.A., Masumoto, H., Larionov, V., Jansen, L.E.T., and Earnshaw, W.C. (2011). Epigenetic engineering shows H3K4me2 is required for

HJURP targeting and CENP-A assembly on a synthetic human kinetochore: H3K4me2 and kinetochore maintenance. *EMBO J.* *30*, 328–340.

Bernard, P., Maure, J.-F., Partridge, J.F., Genier, S., Javerzat, J.-P., and Allshire, R.C. (2001). Requirement of Heterochromatin for Cohesion at Centromeres. *Science* *294*, 2539–2542.

Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., Bouman, P., Liu, J.S., Kouzarides, T., and Schreiber, S.L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci.* *99*, 8695–8700.

Bétermier, M. (2004). Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate *Paramecium*. *Res. Microbiol.* *155*, 399–408.

Bétermier, M., Duharcourt, S., Seitz, H., and Meyer, E. (2000). Timing of Developmentally Programmed Excision and Circularization of *Paramecium* Internal Eliminated Sequences. *Mol. Cell. Biol.* *20*, 1553–1561.

Bianchi, M.E., and Agresti, A. (2005). HMG proteins: dynamic players in gene regulation and differentiation. *Curr. Opin. Genet. Dev.* *15*, 496–506.

Birchler, J.A., Bhadra, M.P., and Bhadra, U. (2000). Making noise about silence: repression of repeated genes in animals. *Curr. Opin. Genet. Dev.* *10*, 211–216.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* *16*, 6–21.

Bird, A.P., and Wolffe, A.P. (1999). Methylation-Induced Repression— Belts, Braces, and Chromatin. *Cell* *99*, 451–454.

Black, B.E., Foltz, D.R., Chakravarthy, S., Luger, K., Woods, V.L., and Cleveland, D.W. (2004). Structural determinants for generating centromeric chromatin. *Nature* *430*, 578–582.

Blower, M.D., and Karpen, G.H. (2001). The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nat. Cell Biol.* *3*, 730–739.

Blower, M.D., Sullivan, B.A., and Karpen, G.H. (2002). Conserved Organization of Centromeric Chromatin in Flies and Humans. *Dev. Cell* *2*, 319–330.

Bouhouche, K., Gout, J.-F., Kapusta, A., Bétermier, M., and Meyer, E. (2011). Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res.* *39*, 4249–4264.

Bouzinba-Segard, H., Guais, A., and Francastel, C. (2006). Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. *Proc. Natl. Acad. Sci.* *103*, 8709–8714.

Brasher, S.V., Smith, B.O., Fogh, R.H., Nietlispach, D., Thiru, A., Nielsen, P.R., Broadhurst, R.W., Ball, L.J., Murzina, N.V., and Laue, E.D. (2000). The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. *EMBO J.* *19*, 1587–1597.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell*

128, 1089–1103.

Buchwald, G., van der Stoep, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M., and Sixma, T.K. (2006). Structure and E3-ligase activity of the Ring–Ring complex of Polycomb proteins Bmi1 and Ring1b. *EMBO J.* *25*, 2465–2474.

Buchwitz, B.J., Ahmad, K., Moore, L.L., Roth, M.B., and Henikoff, S. (1999). Cell division: A histone-H3-like protein in *C. elegans*. *Nature* *401*, 547–548.

Buhler, M., and Gasser, S.M. (2009). Silent chromatin at the middle and ends: lessons from yeasts. *EMBO J* *28*, 2149–2161.

Bui, M., Dimitriadis, E.K., Hoischen, C., An, E., Quénet, D., Giebe, S., Nita-Lazar, A., Diekmann, S., and Dalal, Y. (2012). Cell-Cycle-Dependent Structural Transitions in the Human CENP-A Nucleosome In Vivo. *Cell* *150*, 317–326.

Cao, R., and Zhang, Y. (2004). SUZ12 Is Required for Both the Histone Methyltransferase Activity and the Silencing Function of the EED-EZH2 Complex. *Mol. Cell* *15*, 57–67.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* *298*, 1039–1043.

Castel, S.E., and Martienssen, R.A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.* *14*, 100–112.

Cervantes, M.D., Coyne, R.S., Xi, X., and Yao, M.-C. (2006). The Condensin Complex Is Essential for Amitotic Segregation of Bulk Chromosomes, but Not Nucleoli, in the Ciliate *Tetrahymena thermophila*. *Mol. Cell. Biol.* *26*, 4690–4700.

Chalker, D.L., and Yao, M.-C. (2011). DNA Elimination in Ciliates: Transposon Domestication and Genome Surveillance. *Annu. Rev. Genet.* *45*, 227–246.

Chalker, D.L., and Yao, M.C. (1996). Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of *Tetrahymena thermophila* with germ line-limited DNA. *Mol. Cell. Biol.* *16*, 3658–3667.

Chalker, D.L., Meyer, E., and Mochizuki, K. (2013). Epigenetics of Ciliates. *Cold Spring Harb. Perspect. Biol.* *5*, a017764.

Chan, C.S., Rastelli, L., and Pirrotta, V. (1994). A Polycomb response element in the *Ubx* gene that determines an epigenetically inherited state of repression. *EMBO J.* *13*, 2553.

Cheeseman, I.M., and Desai, A. (2008). Molecular architecture of the kinetochore–microtubule interface. *Nat. Rev. Mol. Cell Biol.* *9*, 33–46.

Cheng, C.-Y., Vogt, A., Mochizuki, K., and Yao, M.-C. (2010). A Domesticated piggyBac Transposase Plays Key Roles in Heterochromatin Dynamics and DNA Cleavage during Programmed DNA Deletion in *Tetrahymena thermophila*. *Mol. Biol. Cell* *21*, 1753–1762.

Ching, A.S.-L., and Berger, J.D. (1986). The timing of initiation of macronuclear DNA synthesis is set

during the preceding cell cycle in *Paramecium tetraurelia*: Analysis of the effects of abrupt changes in nutrient level. *Exp. Cell Res.* 167, 177–190.

Clarke, L., Amstutz, H., Fishel, B., and Carbon, J. (1986). Analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. *Proc. Natl. Acad. Sci.* 83, 8253–8257.

Cléard, F., Moshkin, Y., Karch, F., and Maeda, R.K. (2006). Probing long-distance regulatory interactions in the *Drosophila melanogaster* bithorax complex using Dam identification. *Nat. Genet.* 38, 931–935.

Cmarko, D., Verschure, P.J., Otte, A.P., Driel, R. van, and Fakan, S. (2003). Polycomb group gene silencing proteins are concentrated in the perichromatin compartment of the mammalian nucleus. *J. Cell Sci.* 116, 335–343.

Collins, K.A., Castillo, A.R., Tatsutani, S.Y., and Biggins, S. (2005). De Novo Kinetochores Assembly Requires the Centromeric Histone H3 Variant. *Mol. Biol. Cell* 16, 5649–5660.

Cooper, J.P., Watanabe, Y., and Nurse, P. (1998). Fission yeast Taz1 protein is required for meiotic telomere clustering and recombination. *Nature* 392, 828–831.

Cottarel, G., Shero, J.H., Hieter, P., and Hegemann, J.H. (1989). A 125-base-pair CEN6 DNA fragment is sufficient for complete meiotic and mitotic centromere functions in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 9, 3342.

Coyne, R.S., Lhuillier-Akakpo, M., and Duhaucourt, S. (2012). RNA-guided DNA rearrangements in ciliates: Is the best genome defence a good offence? *Biol. Cell* 104, 309–325.

Di Croce, L., and Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* 20, 1147–1155.

Cui, B., and Gorovsky, M.A. (2006). Centromeric Histone H3 Is Essential for Vegetative Cell Division and for DNA Elimination during Conjugation in *Tetrahymena thermophila*. *Mol. Cell. Biol.* 26, 4499–4510.

Cummings, D.J., Tait, A., and Goddard, J.M. (1974). Methylated bases in DNA from *Paramecium aurelia*. *Biochim. Biophys. Acta* 374, 1–11.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). *Drosophila* Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites. *Cell* 111, 185–196.

Dalal, Y., Wang, H., Lindsay, S., and Henikoff, S. (2007). Tetrameric Structure of Centromeric Nucleosomes in Interphase *Drosophila* Cells. *PLoS Biol* 5, e218.

Déjardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D., and Cavalli, G. (2005). Recruitment of *Drosophila* Polycomb group proteins to chromatin by DSP1. *Nature* 434, 533–538.

Dernburg, A.F. (2001). Here, There, and Everywhere Kinetochores Function on Holocentric Chromosomes. *J. Cell Biol.* 153, F33–F38.

Duhaucourt, S., Butler, A., and Meyer, E. (1995). Epigenetic self-regulation of developmental excision of

an internal eliminated sequence on *Paramecium tetraurelia*. *Genes Dev.* 9, 2065–2077.

Duharcourt, S., Keller, A.-M., and Meyer, E. (1998). Homology-Dependent Maternal Inhibition of Developmental Excision of Internal Eliminated Sequences in *Paramecium tetraurelia*. *Mol. Cell. Biol.* 18, 7075–7085.

Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G. (2009). HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of CENP-A at Centromeres. *Cell* 137, 485–497.

Dunleavy, E.M., Almouzni, G., and Karpen, G.H. (2011). H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G1 phase. *Nucleus* 2, 146–157.

Earnshaw, W.C., and Migeon, B.R. (1985). Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. *Chromosoma* 92, 290–296.

Earnshaw, W.C., and Rothfield, N. (1985). Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* 91, 313–321.

Endoh, M., Endo, T.A., Endoh, T., Isono, K., Sharif, J., Ohara, O., Toyoda, T., Ito, T., Eskeland, R., Bickmore, W.A., et al. (2012). Histone H2A Mono-Ubiquitination Is a Crucial Step to Mediate PRC1-Dependent Repression of Developmental Genes to Maintain ES Cell Identity. *PLoS Genet* 8, e1002774.

Epstein, L.M., and Forney, J.D. (1984). Mendelian and non-mendelian mutations affecting surface antigen expression in *Paramecium tetraurelia*. *Mol. Cell. Biol.* 4, 1583.

Erdel, F., Müller-Ott, K., and Rippe, K. (2013). Establishing epigenetic domains via chromatin-bound histone modifiers. *Ann. N. Y. Acad. Sci.* 1305, 29–43.

Eskeland, R., Eberharter, A., and Imhof, A. (2006). HP1 Binding to Chromatin Methylated at H3K9 Is Enhanced by Auxiliary Factors. *Mol. Cell. Biol.* 27, 453–465.

Filion, G.J., van Bommel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in *Drosophila* Cells. *Cell* 143, 212–224.

Foltz, D.R., Jansen, L.E.T., Black, B.E., Bailey, A.O., Yates, J.R., and Cleveland, D.W. (2006). The human CENP-A centromeric nucleosome-associated complex. *Nat. Cell Biol.* 8, 458–469.

Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates III, J.R., Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. *Cell* 137, 472–484.

Forney, J.D., and Blackburn, E.H. (1988). Developmentally controlled telomere addition in wild-type and mutant paramecia. *Mol. Cell. Biol.* 8, 251.

Francis, N.J., Saurin, A.J., Shao, Z., and Kingston, R.E. (2001). Reconstitution of a Functional Core Polycomb Repressive Complex. *Mol. Cell* 8, 545–556.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin Compaction by a Polycomb Group Protein Complex. *Science* 306, 1574–1577.

- Fujita, Y., Hayashi, T., Kiyomitsu, T., Toyoda, Y., Kokubu, A., Obuse, C., and Yanagida, M. (2007). Priming of Centromere for CENP-A Recruitment by Human hMis18 α , hMis18 β , and M18BP1. *Dev. Cell* 12, 17–30.
- Gao, Z., Fu, S., Dong, Q., Han, F., and Birchler, J.A. (2011). Inactivation of a centromere during the formation of a translocation in maize. *Chromosome Res.* 19, 755–761.
- Garnier, O., Serrano, V., Duharcourt, S., and Meyer, E. (2004). RNA-Mediated Programming of Developmental Genome Rearrangements in *Paramecium tetraurelia*. *Mol. Cell. Biol.* 24, 7370–7379.
- Ge, D.T., and Zamore, P.D. (2013). Small RNA-directed silencing: the fly finds its inner fission yeast? *Curr. Biol. CB* 23, R318–320.
- Gindullis, F., Desel, C., Galasso, I., and Schmidt, T. (2001). The Large-Scale Organization of the Centromeric Region in Beta Species. *Genome Res.* 11, 253–265.
- Gratias, A., and Bétermier, M. (2003). Processing of Double-Strand Breaks Is Involved in the Precise Excision of *Paramecium* Internal Eliminated Sequences. *Mol. Cell. Biol.* 23, 7152–7162.
- Gratias, A., and Bétermier, M. (2001). Developmentally programmed excision of internal DNA sequences in *Paramecium aurelia*. *Biochimie* 83, 1009–1022.
- Gratias, A., Lepère, G., Garnier, O., Rosa, S., Duharcourt, S., Malinsky, S., Meyer, E., and Bétermier, M. (2008). Developmentally programmed DNA splicing in *Paramecium* reveals short-distance crosstalk between DNA cleavage sites. *Nucleic Acids Res.* 36, 3244–3251.
- Grau, D.J., Chapman, B.A., Garlick, J.D., Borowsky, M., Francis, N.J., and Kingston, R.E. (2011). Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev.* 25, 2210–2221.
- Grewal, S.I. (2010). RNAi-dependent formation of heterochromatin and its diverse functions. *Curr. Opin. Genet. Dev.* 20, 134–141.
- Grewal, S.I.S., and Jia, S. (2007). Heterochromatin revisited. *Nat. Rev. Genet.* 8, 35–46.
- Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006). RNAi Components Are Required for Nuclear Clustering of Polycomb Group Response Elements. *Cell* 124, 957–971.
- Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Dénervaud, N., Bucher, P., and Trono, D. (2010). KRAB–Zinc Finger Proteins and KAP1 Can Mediate Long-Range Transcriptional Repression through Heterochromatin Spreading. *PLoS Genet* 6, e1000869.
- Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5' End Formation in *Drosophila*. *Science* 315, 1587–1590.
- Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076.

- Hall, I.M., Noma, K., and Grewal, S.I.S. (2003). RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc. Natl. Acad. Sci.* *100*, 193–198.
- Han, F., Lamb, J.C., and Birchler, J.A. (2006). High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 3238–3243.
- Han, F., Gao, Z., and Birchler, J.A. (2009). Reactivation of an Inactive Centromere Reveals Epigenetic and Structural Components for Centromere Specification in Maize. *Plant Cell Online* *21*, 1929–1939.
- Harrison, G.S., and Karrer, K.M. (1985). DNA synthesis, methylation and degradation during conjugation in *Tetrahymena thermophila*. *Nucleic Acids Res.* *13*, 73–87.
- Hayashi, T., Fujita, Y., Iwasaki, O., Adachi, Y., Takahashi, K., and Yanagida, M. (2004). Mis16 and Mis18 Are Required for CENP-A Loading and Histone Deacetylation at Centromeres. *Cell* *118*, 715–729.
- Heeger, S., Leismann, O., Schittenhelm, R., Schraidt, O., Heidmann, S., and Lehner, C.F. (2005). Genetic interactions of separate regulatory subunits reveal the diverged *Drosophila* Cenp-C homolog. *Genes Dev.* *19*, 2041–2053.
- Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* *39*, 311–318.
- Hekimoglu, B., and Ringrose, L. (2009). Non-coding RNAs in Polycomb/Trithorax regulation. *RNA Biol.* *6*, 129–137.
- Henikoff, S., and Dalal, Y. (2005). Centromeric chromatin: what makes it unique? *Curr. Opin. Genet. Dev.* *15*, 177–184.
- Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D., and Karpen, G.H. (2006). Mislocalization of the *Drosophila* Centromere-Specific Histone CID Promotes Formation of Functional Ectopic Kinetochores. *Dev. Cell* *10*, 303–315.
- Higgins, A.W., Schueler, M.G., and Willard, H.F. (1999). Chromosome engineering: generation of mono- and dicentric isochromosomes in a somatic cell hybrid system. *Chromosoma* *108*, 256–265.
- Higgins, A.W., Gustashaw, K.M., and Willard, H.F. (2005). Engineered human dicentric chromosomes show centromere plasticity. *Chromosome Res.* *13*, 745–762.
- Hill, A., and Bloom, K. (1989). Acquisition and processing of a conditional dicentric chromosome in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *9*, 1368.
- Hodgson, J.W., Argiropoulos, B., and Brock, H.W. (2001). Site-Specific Recognition of a 70-Base-Pair Element Containing d(GA)_n Repeats Mediates bithoraxoid Polycomb Group Response Element-Dependent Silencing. *Mol. Cell. Biol.* *21*, 4528–4543.
- Horard, B., Tatout, C., Poux, S., and Pirrotta, V. (2000). Structure of a Polycomb Response Element and In Vitro Binding of Polycomb Group Complexes Containing GAGA Factor. *Mol. Cell. Biol.* *20*, 3187.
- Hori, T., Haraguchi, T., Hiraoka, Y., Kimura, H., and Fukagawa, T. (2003). Dynamic behavior of Nuf2-

Hec1 complex that localizes to the centrosome and centromere and is essential for mitotic progression in vertebrate cells. *J. Cell Sci.* 116, 3347–3362.

Howman, E.V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P., and Choo, K.H.A. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc. Natl. Acad. Sci.* 97, 1148–1153.

Hu, H., Liu, Y., Wang, M., Fang, J., Huang, H., Yang, N., Li, Y., Wang, J., Yao, X., Shi, Y., et al. (2011). Structure of a CENP-A–histone H4 heterodimer in complex with chaperone HJURP. *Genes Dev.* 25, 901–906.

Huang, J., Dorsey, J., Chuikov, S., Zhang, X., Jenuwein, T., Reinberg, D., and Berger, S.L. (2010). G9a and Glp Methylate Lysine 373 in the Tumor Suppressor p53. *J. Biol. Chem.* 285, 9636–9641.

Jackson, J.P., Johnson, L., Jasencakova, Z., Zhang, X., PerezBurgos, L., Singh, P.B., Cheng, X., Schubert, I., Jenuwein, T., and Jacobsen, S.E. (2004). Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in *Arabidopsis thaliana*. *Chromosoma* 112, 308–315.

Jacobs, S.A., Taverna, S.D., Zhang, Y., Briggs, S.D., Li, J., Eissenberg, J.C., Allis, C.D., and Khorasanizadeh, S. (2001). Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J.* 20, 5232–5241.

Jansen, L.E.T., Black, B.E., Foltz, D.R., and Cleveland, D.W. (2007). Propagation of centromeric chromatin requires exit from mitosis. *J. Cell Biol.* 176, 795–805.

Jia, S., Noma, K., and Grewal, S.I.S. (2004). RNAi-Independent Heterochromatin Nucleation by the Stress-Activated ATF/CREB Family Proteins. *Science* 304, 1971–1976.

Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., and Jacobsen, S.E. (2007). The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation. *Curr. Biol.* 17, 379–384.

Kagansky, A., Folco, H.D., Almeida, R., Pidoux, A.L., Boukaba, A., Simmer, F., Urano, T., Hamilton, G.L., and Allshire, R.C. (2009). Synthetic Heterochromatin Bypasses RNAi and Centromeric Repeats to Establish Functional Centromeres. *Science* 324, 1716–1719.

Kanoh, J., Sadaie, M., Urano, T., and Ishikawa, F. (2005). Telomere Binding Protein Taz1 Establishes Swi6 Heterochromatin Independently of RNAi at Telomeres. *Curr. Biol.* 15, 1808–1819.

Kapusta, A., Matsuda, A., Marmignon, A., Ku, M., Silve, A., Meyer, E., Forney, J.D., Malinsky, S., and Bétermier, M. (2011). Highly Precise and Developmentally Programmed Genome Assembly in *Paramecium* Requires Ligase IV–Dependent End Joining. *PLoS Genet* 7, e1002049.

Karagianni, P., Amazit, L., Qin, J., and Wong, J. (2008). ICBP90, a Novel Methyl K9 H3 Binding Protein Linking Protein Ubiquitination with Heterochromatin Formation. *Mol. Cell. Biol.* 28, 705–717.

Ketel, C.S., Andersen, E.F., Vargas, M.L., Suh, J., Strome, S., and Simon, J.A. (2005). Subunit Contributions to Histone Methyltransferase Activities of Fly and Worm Polycomb Group Complexes. *Mol. Cell. Biol.* 25, 6857–6868.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J.,

- Larschan, E., Gorchakov, A.A., Gu, T., et al. (2011). Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* *471*, 480–485.
- Kim, H., Kang, K., and Kim, J. (2009). AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res.* *37*, 2940–2950.
- King, I.F.G., Francis, N.J., and Kingston, R.E. (2002). Native and Recombinant Polycomb Group Complexes Establish a Selective Block to Template Accessibility To Repress Transcription In Vitro. *Mol. Cell. Biol.* *22*, 7919–7928.
- Klobutcher, L.A., and Herrick, G. (1995). Consensus inverted terminal repeat sequence of *Paramecium* IESs: resemblance to termini of Tc1-related and *Euplotes* Tec transposons. *Nucleic Acids Res.* *23*, 2006.
- Klobutcher, L.A., and Herrick, G. (1997). Developmental genome reorganization in ciliated protozoa: the transposon link. *Prog. Nucleic Acid Res. Mol. Biol.* *56*, 1–62.
- Koo, D.-H., Han, F., Birchler, J.A., and Jiang, J. (2011). Distinct DNA methylation patterns associated with active and inactive centromeres of the maize B chromosome. *Genome Res.* *21*, 908–914.
- Koshland, D., Rutledge, L., Fitzgerald-Hayes, M., and Hartwell, L.H. (1987). A genetic analysis of dicentric minichromosomes in *saccharomyces cerevisiae*. *Cell* *48*, 801–812.
- Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., and Xiong, Y. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15INK4B tumor suppressor gene. *Oncogene* *30*, 1956–1962.
- Kurth, H.M., and Mochizuki, K. (2009). 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in *Tetrahymena*. *RNA* *15*, 675–685.
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* *16*, 2893–2905.
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* *410*, 116–120.
- Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jørgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., et al. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat. Cell Biol.* *12*, 618–624.
- Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* *9*, 1167–1174.
- Lee, S.R., and Collins, K. (2006). Two classes of endogenous small RNAs in *Tetrahymena thermophila*. *Genes Dev.* *20*, 28–33.
- Lee, J.M., Lee, J.S., Kim, H., Kim, K., Park, H., Kim, J.-Y., Lee, S.H., Kim, I.S., Kim, J., Lee, M., et al. (2012). EZH2 Generates a Methyl Degron that Is Recognized by the DCAF1/DDB1/CUL4 E3 Ubiquitin Ligase Complex. *Mol. Cell* *48*, 572–586.

- Lehmann, L., Ferrari, R., Vashisht, A.A., Wohlschlegel, J.A., Kurdiani, S.K., and Carey, M. (2012). Polycomb Repressive Complex 1 (PRC1) Disassembles RNA Polymerase II Preinitiation Complexes. *J. Biol. Chem.* 287, 35784–35794.
- Lepère, G., Bétermier, M., Meyer, E., and Duharcourt, S. (2008). Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev.* 22, 1501–1512.
- Lepère, G., Nowacki, M., Serrano, V., Gout, J.-F., Guglielmi, G., Duharcourt, S., and Meyer, E. (2009). Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res.* 37, 903–915.
- Li, L., and Liu, Y. (2011). Diverse Small Non-coding RNAs in RNA Interference Pathways. In *Therapeutic Oligonucleotides*, J. Goodchild, ed. (Humana Press), pp. 169–182.
- Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* 24, 368–380.
- Litt, M.D., Simpson, M., Gaszner, M., Allis, C.D., and Felsenfeld, G. (2001). Correlation Between Histone Lysine Methylation and Developmental Changes at the Chicken β -Globin Locus. *Science* 293, 2453–2455.
- Liu, S.-T., Hittle, J.C., Jablonski, S.A., Campbell, M.S., Yoda, K., and Yen, T.J. (2003). Human CENP-I specifies localization of CENP-F, MAD1 and MAD2 to kinetochores and is essential for mitosis. *Nat. Cell Biol.* 5, 341–345.
- Liu, S.-T., Rattner, J.B., Jablonski, S.A., and Yen, T.J. (2006). Mapping the assembly pathways that specify formation of the trilaminar kinetochore plates in human cells. *J. Cell Biol.* 175, 41–53.
- Liu, Y., Mochizuki, K., and Gorovsky, M.A. (2004). Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in *Tetrahymena*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1679.
- Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in *Tetrahymena*. *Genes Dev.* 21, 1530–1545.
- Lomberk, G., Wallrath, L., and Urrutia, R. (2006). The Heterochromatin Protein 1 family. *Genome Biol.* 7, 228.
- Maddox, P.S., Hyndman, F., Monen, J., Oegema, K., and Desai, A. (2007). Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. *J. Cell Biol.* 176, 757–763.
- Madireddi, M.T., Coyne, R.S., Smothers, J.F., Mickey, K.M., Yao, M.-C., and Allis, C.D. (1996). Pdd1p, A Novel Chromodomain-Containing Protein, Links Heterochromatin Assembly and DNA Elimination in *Tetrahymena*. *Cell* 87, 75–84.
- Malone, C.D., Anderson, A.M., Motl, J.A., Rexer, C.H., and Chalker, D.L. (2005). Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of *Tetrahymena thermophila*. *Mol. Cell. Biol.* 25, 9151.

- Mansfield, R.E., Musselman, C.A., Kwan, A.H., Oliver, S.S., Garske, A.L., Davrazou, F., Denu, J.M., Kutateladze, T.G., and Mackay, J.P. (2011). Plant Homeodomain (PHD) Fingers of CHD4 Are Histone H3-binding Modules with Preference for Unmodified H3K4 and Methylated H3K9. *J. Biol. Chem.* *286*, 11779–11791.
- Mao, Y.S., Zhang, B., and Spector, D.L. (2011). Biogenesis and function of nuclear bodies. *Trends Genet.* *27*, 295–306.
- Matthews, A.G.W., Kuo, A.J., Ramon-Maiques, S., Han, S., Champagne, K.S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D.N., et al. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* *450*, 1106–1110.
- May, B.P., Lippman, Z.B., Fang, Y., Spector, D.L., and Martienssen, R.A. (2005). Differential Regulation of Strand-Specific Transcripts from Arabidopsis Centromeric Satellite Repeats. *PLoS Genet* *1*, e79.
- Mayer, K.M., and Forney, J.D. (1999). A Mutation in the Flanking 5'-TA-3' Dinucleotide Prevents Excision of an Internal Eliminated Sequence From the Paramecium tetraurelia Genome. *Genetics* *151*, 597–604.
- Mayer, K.M., Mikami, K., and Forney, J.D. (1998). A mutation in Paramecium tetraurelia reveals functional and structural features of developmentally excised DNA elements. *Genetics* *148*, 139–149.
- McClintock, B. (1939). The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis. *Proc. Natl. Acad. Sci. U. S. A.* *25*, 405.
- McClintock, B. (1941). The Stability of Broken Ends of Chromosomes in Zea Mays. *Genetics* *26*, 234–282.
- Melters, D.P., Paliulis, L.V., Korf, I.F., and Chan, S.W.L. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* *20*, 579–593.
- Meyer, E., and Garnier, O. (2002). Non-Mendelian inheritance and homology-dependent effects in ciliates. *Adv. Genet.* *46*, 305–337.
- Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., and Lis, J.T. (2011). Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev.* *25*, 742–754.
- Mitchell, A.R., Gosden, J.R., and Miller, D.A. (1985). A cloned sequence, p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. *Chromosoma* *92*, 369–377.
- Mitra, R., Fain-Thornton, J., and Craig, N.L. (2008). piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J.* *27*, 1097–1109.
- Mizuguchi, G., Xiao, H., Wisniewski, J., Smith, M.M., and Wu, C. (2007). Nonhistone Scm3 and Histones CenH3-H4 Assemble the Core of Centromere-Specific Nucleosomes. *Cell* *129*, 1153–1164.
- Mochizuki, K., and Gorovsky, M.A. (2004). Small RNAs in genome rearrangement in Tetrahymena. *Curr. Opin. Genet. Dev.* *14*, 181–187.
- Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-Related Gene

Implicates Small RNAs in Genome Rearrangement in Tetrahymena. *Cell* 110, 689–699.

Le Mouel, A., Butler, A., Caron, F., and Meyer, E. (2003). Developmentally Regulated Chromosome Fragmentation Linked to Imprecise Elimination of Repeated Sequences in Paramecia. *Eukaryot. Cell* 2, 1076–1090.

Müller, J., and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr. Opin. Genet. Dev.* 16, 476–484.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone Methyltransferase Activity of a *Drosophila* Polycomb Group Repressor Complex. *Cell* 111, 197–208.

Murata, M., Ogura, Y., and Motoyoshi, F. (1994). Centromeric repetitive sequences in *Arabidopsis thaliana*. *Idengaku Zasshi* 69, 361–370.

Musselman, C.A., Mansfield, R.E., Garske, A.L., Davrazou, F., Kwan, A.H., Oliver, S.S., O'Leary, H., Denu, J.M., Mackay, J.P., and Kutateladze, T.G. (2009). Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem. J.* 423, 179–187.

Nakaseko, Y., Adachi, Y., Funahashi, S., Niwa, O., and Yanagida, M. (1986). Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast. *EMBO J.* 5, 1011.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I.S. (2001). Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. *Science* 292, 110–113.

Neumann, P., Yan, H., and Jiang, J. (2007). The Centromeric Retrotransposons of Rice Are Transcribed and Differentially Processed by RNA Interference. *Genetics* 176, 749–761.

Noto, T., Kurth, H.M., Kataoka, K., Aronica, L., DeSouza, L.V., Siu, K.W.M., Pearlman, R.E., Gorovsky, M.A., and Mochizuki, K. (2010). The Tetrahymena Argonaute-Binding Protein Giw1p Directs a Mature Argonaute-siRNA Complex to the Nucleus. *Cell* 140, 692–703.

Nowacki, M., Zagorski-Ostojka, W., and Meyer, E. (2005). Nowa1p and Nowa2p: Novel Putative RNA Binding Proteins Involved in trans-Nuclear Crosstalk in *Paramecium tetraurelia*. *Curr. Biol.* 15, 1616–1628.

Nowak, J.K., Gromadka, R., Juszczuk, M., Jerka-Dziadosz, M., Maliszewska, K., Mucchielli, M.-H., Gout, J.-F., Arnaiz, O., Agier, N., Tang, T., et al. (2011). Functional study of genes essential for autogamy and nuclear reorganization in *Paramecium*. *Eukaryot. Cell* 10, 363–372.

O'Geen, H., Squazzo, S.L., Iyengar, S., Blahnik, K., Rinn, J.L., Chang, H.Y., Green, R., and Farnham, P.J. (2007). Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. *PLoS Genet.* 3, e89.

Oegema, K., Desai, A., Rybina, S., Kirkham, M., and Hyman, A.A. (2001). Functional Analysis of Kinetochores Assembly in *Caenorhabditis elegans*. *J. Cell Biol.* 153, 1209–1226.

Ohzeki, J., Nakano, M., Okada, T., and Masumoto, H. (2002). CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. *J. Cell Biol.* 159, 765–775.

- Ohzeki, J., Bergmann, J.H., Kouprina, N., Noskov, V.N., Nakano, M., Kimura, H., Earnshaw, W.C., Larionov, V., and Masumoto, H. (2012). Breaking the HAC Barrier: Histone H3K9 acetyl/methyl balance regulates CENP-A assembly: Breaking the HAC Barrier. *EMBO J.* *31*, 2391–2402.
- Page, S.L., and Shaffer, L.G. (1998). Chromosome stability is maintained by short intercentromeric distance in functionally dicentric human Robertsonian translocations. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* *6*, 115–122.
- Parker, J.S., and Barford, D. (2006). Argonaute: a scaffold for the function of short regulatory RNAs. *Trends Biochem. Sci.* *31*, 622–630.
- Pasini, D., Cloos, P.A.C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* *464*, 306–310.
- Pawlowski, J. (2013). The new micro-kingdoms of eukaryotes. *BMC Biol.* *11*, 40.
- Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells. *Cell* *139*, 1290–1302.
- Pennaneach, V., and Kolodner, R.D. (2009). Stabilization of Dicentric Translocations through Secondary Rearrangements Mediated by Multiple Mechanisms in *S. cerevisiae*. *PLoS ONE* *4*, e6389.
- Pirrotta, V., and Li, H.-B. (2012). A view of nuclear Polycomb bodies. *Curr. Opin. Genet. Dev.* *22*, 101–109.
- Plasterk, R.H., Izsvák, Z., and Ivics, Z. (1999). Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet. TIG* *15*, 326–332.
- Ranjan, A., Mizuguchi, G., FitzGerald, P.C., Wei, D., Wang, F., Huang, Y., Luk, E., Woodcock, C.L., and Wu, C. (2013). Nucleosome-free Region Dominates Histone Acetylation in Targeting SWR1 to Promoters for H2A.Z Replacement. *Cell* *154*, 1232–1245.
- Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* *406*, 593–599.
- Rice, J.C., Briggs, S.D., Ueberheide, B., Barber, C.M., Shabanowitz, J., Hunt, D.F., Shinkai, Y., and Allis, C.D. (2003). Histone Methyltransferases Direct Different Degrees of Methylation to Define Distinct Chromatin Domains. *Mol. Cell* *12*, 1591–1598.
- Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* *129*, 1311–1323.
- Round, E.K., Flowers, S.K., and Richards, E.J. (1997). *Arabidopsis thaliana* Centromere Regions: Genetic Map Positions and Repetitive DNA Structure. *Genome Res.* *7*, 1045–1053.
- Saffery, R., Irvine, D.V., Griffiths, B., Kalitsis, P., Wordeman, L., and Choo, K.H.A. (2000). Human centromeres and neocentromeres show identical distribution patterns of >20 functionally important

kinetochore-associated proteins. *Hum. Mol. Genet.* **9**, 175–185.

Saffery, R., Sumer, H., Hassan, S., Wong, L.H., Craig, J.M., Todokoro, K., Anderson, M., Stafford, A., and Choo, K.H.A. (2003). Transcription within a Functional Human Centromere. *Mol. Cell* **12**, 509–516.

Saitoh, H., Tomkiel, J., Cooke, C.A., Ratrie III, H., Maurer, M., Rothfield, N.F., and Earnshaw, W.C. (1992). CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. *Cell* **70**, 115–125.

Sandoval, P.Y., Swart, E.C., Arambasic, M., and Nowacki, M. (2014). Functional Diversification of Dicer-like Proteins and Small RNAs Required for Genome Sculpting. *Dev. Cell* **28**, 174–188.

Sanyal, K., Baum, M., and Carbon, J. (2004). Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are all different and unique. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11374–11379.

Sarma, K., and Reinberg, D. (2005). Histone variants meet their match. *Nat. Rev. Mol. Cell Biol.* **6**, 139–149.

Sato, H., Masuda, F., Takayama, Y., Takahashi, K., and Saitoh, S. (2012). Epigenetic Inactivation and Subsequent Heterochromatinization of a Centromere Stabilize Dicentric Chromosomes. *Curr. Biol.* **22**, 658–667.

Schotta, G. (2002). Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J.* **21**, 1121–1131.

Schuh, M., Lehner, C.F., and Heidmann, S. (2007). Incorporation of *Drosophila* CID/CENP-A and CENP-C into Centromeres during Early Embryonic Anaphase. *Curr. Biol.* **17**, 237–243.

Schwartz, Y.B., and Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.* **14**, 853–864.

Seum, C., Delattre, M., Spierer, A., and Spierer, P. (2001). Ectopic HP1 promotes chromosome loops and variegated silencing in *Drosophila*. *EMBO J.* **20**, 812–818.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C., Bender, W., and Kingston, R.E. (1999). Stabilization of Chromatin Structure by PRC1, a Polycomb Complex. *Cell* **98**, 37–46.

Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.-C., Lee, Y., and Orkin, S.H. (2009). Jumonji Modulates Polycomb Activity and Self-Renewal versus Differentiation of Stem Cells. *Cell* **139**, 1303–1314.

Shivaraju, M., Unruh, J.R., Slaughter, B.D., Mattingly, M., Berman, J., and Gerton, J.L. (2012). Cell-Cycle-Coupled Structural Oscillation of Centromeric Nucleosomes in Yeast. *Cell* **150**, 304–316.

Shuaib, M., Ouarrhni, K., Dimitrov, S., and Hamiche, A. (2010). HJURP binds CENP-A via a highly conserved N-terminal domain and mediates its deposition at centromeres. *Proc. Natl. Acad. Sci.* **107**, 1349–1354.

Simon, J.A., and Kingston, R.E. (2013). Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Mol. Cell* **49**, 808–824.

- Smith, S., and Stillman, B. (1991). Stepwise assembly of chromatin during DNA replication in vitro. *EMBO J.* *10*, 971–980.
- Sonneborn, T.M. (1977). Genetics of Cellular Differentiation: Stable Nuclear Differentiation in Eucaryotic Unicells. *Annu. Rev. Genet.* *11*, 349–367.
- Spencer, F., and Hieter, P. (1992). Centromere DNA mutations induce a mitotic delay in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* *89*, 8908–8912.
- Sripathy, S.P., Stevens, J., and Schultz, D.C. (2006). The KAP1 Corepressor Functions To Coordinate the Assembly of De Novo HP1-Demarcated Microenvironments of Heterochromatin Required for KRAB Zinc Finger Protein-Mediated Transcriptional Repression. *Mol. Cell. Biol.* *26*, 8623–8638.
- Steele, C.J., Barkocy-Gallagher, G.A., Preer, L.B., and Preer, J.R. (1994). Developmentally excised sequences in micronuclear DNA of *Paramecium*. *Proc. Natl. Acad. Sci.* *91*, 2255–2259.
- Stimpson, K.M., and Sullivan, B.A. (2010). Epigenomics of centromere assembly and function. *Curr. Opin. Cell Biol.* *22*, 772–780.
- Stimpson, K.M., Song, I.Y., Jauch, A., Holtgreve-Grez, H., Hayden, K.E., Bridger, J.M., and Sullivan, B.A. (2010). Telomere Disruption Results in Non-Random Formation of De Novo Dicentric Chromosomes Involving Acrocentric Human Chromosomes. *PLoS Genet* *6*, e1001061.
- Stimpson, K.M., Matheny, J.E., and Sullivan, B.A. (2012). Dicentric chromosomes: unique models to study centromere function and inactivation. *Chromosome Res.* *20*, 595–605.
- Stoler, S., Keith, K.C., Curnick, K.E., and Fitzgerald-Hayes, M. (1995). A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. *Genes Dev.* *9*, 573–586.
- Sullivan, B.A., and Karpen, G.H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat. Struct. Mol. Biol.* *11*, 1076–1083.
- Sullivan, B.A., and Schwartz, S. (1995). Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. *Hum. Mol. Genet.* *4*, 2189–2197.
- Sullivan, B.A., and Willard, H.F. (1998). Stable dicentric X chromosomes with two functional centromeres. *Nat. Genet.* *20*, 227–228.
- Sullivan, B.A., Wolff, D.J., and Schwartz, S. (1994a). Analysis of centromeric activity in Robertsonian translocations: implications for a functional acrocentric hierarchy. *Chromosoma* *103*, 459–467.
- Sullivan, B.A., Blower, M.D., and Karpen, G.H. (2001). Determining centromere identity: cyclical stories and forking paths. *Nat. Rev. Genet.* *2*, 584–596.
- Sullivan, K.F., Hechenberger, M., and Masri, K. (1994b). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J. Cell Biol.* *127*, 581–592.
- Sun, X., Le, H.D., Wahlstrom, J.M., and Karpen, G.H. (2003). Sequence Analysis of a Functional *Drosophila* Centromere. *Genome Res.* *13*, 182–194.

- Szerlong, H.J., and Hansen, J.C. (2011). Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochem. Cell Biol. Biochim. Biol. Cell.* 89, 24–34.
- Takahashi, K., Chen, E.S., and Yanagida, M. (2000). Requirement of Mis6 Centromere Connector for Localizing a CENP-A-Like Protein in Fission Yeast. *Science* 288, 2215–2219.
- Taverna, S.D., Coyne, R.S., and Allis, C.D. (2002). Methylation of Histone H3 at Lysine 9 Targets Programmed DNA Elimination in *Tetrahymena*. *Cell* 110, 701–711.
- Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science* 329, 689–693.
- Tucker, J.B., Beisson, J., Roche, D.L., and Cohen, J. (1980). Microtubules and control of macronuclear “amitosis” in *Paramecium*. *J. Cell Sci.* 44, 135–151.
- Vermaak, D., Hayden, H.S., and Henikoff, S. (2002). Centromere Targeting Element within the Histone Fold Domain of Cid. *Mol. Cell. Biol.* 22, 7553–7561.
- Verschure, P.J., van der Kraan, I., de Leeuw, W., van der Vlag, J., Carpenter, A.E., Belmont, A.S., and van Driel, R. (2005). In Vivo HP1 Targeting Causes Large-Scale Chromatin Condensation and Enhanced Histone Lysine Methylation. *Mol. Cell. Biol.* 25, 4552–4564.
- Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.-M., et al. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439, 871–874.
- Vogel, M.J., Guelen, L., Wit, E. de, Hupkes, D.P., Lodén, M., Talhout, W., Feenstra, M., Abbas, B., Classen, A.-K., and Steensel, B. van (2006). Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res.* 16, 1493–1504.
- Vogt, A., and Mochizuki, K. (2013). A Domesticated PiggyBac Transposase Interacts with Heterochromatin and Catalyzes Reproducible DNA Elimination in *Tetrahymena*. *PLoS Genet* 9, e1004032.
- Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. *Science* 297, 1833–1837.
- Warburton, P.E. (2004). Chromosomal dynamics of human neocentromere formation. *Chromosome Res.* 12, 617–626.
- Weaver, B.A.A., and Cleveland, D.W. (2007). Aneuploidy: Instigator and Inhibitor of Tumorigenesis. *Cancer Res.* 67, 10103–10105.
- Westermann, S., Drubin, D.G., and Barnes, G. (2007). Structures and Functions of Yeast Kinetochore Complexes. *Annu. Rev. Biochem.* 76, 563–591.
- Willard, H.F. (1985). Chromosome-specific organization of human alpha satellite DNA. *Am. J. Hum. Genet.* 37, 524.
- Williams, B.C., Murphy, T.D., Goldberg, M.L., and Karpen, G.H. (1998). Neocentromere activity of

structurally acentric mini-chromosomes in *Drosophila*. *Nat. Genet.* *18*, 30–38.

Yamada, T., Fischle, W., Sugiyama, T., Allis, C.D., and Grewal, S.I.S. (2005). The Nucleation and Maintenance of Heterochromatin by a Histone Deacetylase in Fission Yeast. *Mol. Cell* *20*, 173–185.

Yan, H., Ito, H., Nobuta, K., Ouyang, S., Jin, W., Tian, S., Lu, C., Venu, R.C., Wang, G., Green, P.J., et al. (2006). Genomic and Genetic Characterization of Rice Cen3 Reveals Extensive Transcription and Evolutionary Implications of a Complex Centromere. *Plant Cell Online* *18*, 2123–2133.

Yao, M.-C., Choi, J., Yokoyama, S., Austerberry, C.F., and Yao, C.-H. (1984). DNA elimination in tetrahymena: A developmental process involving extensive breakage and rejoining of DNA at defined sites. *Cell* *36*, 433–440.

Yap, K.L., Li, S., Muñoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.-M. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell* *38*, 662–674.

Yasuhara, J.C., and Wakimoto, B.T. (2006). Oxymoron no more: the expanding world of heterochromatic genes. *Trends Genet.* *22*, 330–338.

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., et al. (2012). Dense Chromatin Activates Polycomb Repressive Complex 2 to Regulate H3 Lysine 27 Methylation. *Science* *337*, 971–975.

Zagulski, M., Nowak, J.K., Le Mouél, A., Nowacki, M., Migdalski, A., Gromadka, R., Noël, B., Blanc, I., Dessen, P., Wincker, P., et al. (2004). High Coding Density on the Largest *Paramecium tetraurelia* Somatic Chromosome. *Curr. Biol.* *14*, 1397–1404.

Zhang, W., Friebe, B., Gill, B.S., and Jiang, J. (2010). Centromere inactivation and epigenetic modifications of a plant chromosome with three functional centromeres. *Chromosoma* *119*, 553–563.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. *Science* *322*, 750–756.

ANNEXE 1

Genome-defence small RNAs exapted for epigenetic mating-type inheritance

Deepankar Pratap Singh^{1,2}, Baptiste Saudemont^{1,2,9}, Gérard Guglielmi¹, Olivier Arnaiz³, Jean-François Gôut^{4,10}, Malgorzata Prajer⁵, Alexey Potekhin⁶, Ewa Przybòs⁵, Anne Aubusson-Fleury³, Simran Bhullar¹, Khaled Bouhouche^{1,11}, Maoussi Lhuillier-Akakpo^{2,7}, Véronique Tanty¹, Corinne Blugeon¹, Adriana Alberti⁸, Karine Labadie⁸, Jean-Marc Aury⁸, Linda Sperling³, Sandra Duharcourt⁷ & Eric Meyer¹

¹Ecole Normale Supérieure, Institut de Biologie de l'ENS, IBENS; Inserm, U1024; CNRS, UMR 8197; Paris, F-75005 France.

²Sorbonne Universités, UPMC Univ., IFD, 4 place Jussieu, 75252 Paris cedex05, France.

³CNRS UPR3404 Centre de Génétique Moléculaire, Gif-sur-Yvette F-91198 cedex, France; Université Paris-Sud, Département de Biologie, Orsay, F-91405, France.

⁴CNRS UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon, 43 boulevard du 11 Novembre 1918, Villeurbanne, F-69622, France.

⁵Institute of Systematics and Evolution of Animals, Polish Academy of Sciences, Ślawkowska 17, 31-016 Krakow, Poland.

⁶Dept of Microbiology, Faculty of Biology and Soil Science, St. Petersburg State University, Saint Petersburg 199034, Russia.

⁷Institut Jacques Monod, CNRS, UMR 7592, Université Paris Diderot, Sorbonne Paris Cité, Paris, F-75005 France.

⁸Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Genoscope, 2 rue Gaston Crémieux, BP5706, 91057 Evry, France.

Present affiliations:

⁹Laboratoire de Biochimie, Unité Mixte de Recherche 8231, École Supérieure de Physique et de Chimie Industrielles, 75231 Paris, France.

¹⁰Department of Biology, Indiana University, Bloomington, Indiana 47405, USA.

¹¹INRA, UMR 1061 Unité de Génétique Moléculaire Animale, Université de Limoges, IFR 145, Faculté des Sciences et Techniques, 87060 Limoges, France.

Correspondence should be addressed to E.M. (emeyer@biologie.ens.fr). Institut de Biologie de l'ENS, IBENS, 46 rue d'Ulm, 75005 Paris, France.

Abstract

In the ciliate *Paramecium*, transposable elements and their single-copy remnants are deleted during the development of somatic macronuclei from germline micronuclei, at each sexual generation. Deletions are targeted by scnRNAs, small RNAs produced from the germline during meiosis that first scan the maternal macronuclear genome to identify missing sequences, and then allow the zygotic macronucleus to reproduce the same deletions. Here, we show that this process accounts for the maternal inheritance of mating types in *P. tetraurelia*, a long-standing problem in epigenetics. Mating type E depends on expression of the transmembrane protein mtA, and the default type O is determined during development by scnRNA-dependent excision of the *mtA* promoter. In the sibling species *P. septaurelia*, mating type O is determined by coding-sequence deletions in a different gene, *mtB*, which is specifically required for *mtA* expression. These independently evolved mechanisms suggest frequent exaptation of the scnRNA pathway to regulate cellular genes and mediate transgenerational epigenetic inheritance of essential phenotypic polymorphisms.

Introduction

Ciliates are complex unicellular eukaryotes that use different types of nuclei within the same cytoplasm to separate germline and somatic functions¹. The diploid micronuclei (MICs) undergo meiosis to provide gametic nuclei during sexual events, but their genome is not expressed. Genes are expressed from the polyploid macronucleus (MAC), which is not transmitted across sexual generations. Following meiosis and fertilization, the parental MAC is lost and replaced by a new one that develops from a mitotic copy of the zygotic nucleus. Macronuclear development involves extensive rearrangements of the germline genome, including the elimination of virtually all Transposable Elements (TEs) and other repeats². Furthermore, in *P. tetraurelia* ~45,000 short, single-copy Internal Eliminated Sequences (IESs) are precisely excised from coding and non-coding sequences^{3,4} by the domesticated transposase Pgm⁵. IESs are invariably flanked by two 5'-TA-3' dinucleotides which recombine into one after excision⁴. A short consensus adjacent to the TAs (5'-TAYAGYNR-3') is reminiscent of the ends of Tc1/mariner elements^{4,6}, and a recent study provided support for the hypothesis that IESs are degenerate remnants of ancient TE insertions^{3,7}.

All intragenic IESs identified so far must be excised to reconstitute functional genes in the MAC. However, the poorly conserved IES end consensus is not sufficient to specify the excision pattern genome-wide⁴, and a specific class of small RNAs is required to identify some IESs on the basis of their absence from the parental MAC⁸⁻¹⁰. In the current “genome scanning” model, scnRNAs are produced from most of the germline genome during MIC meiosis^{11,12} and are then filtered by pairing interactions with nascent transcripts in the parental MAC¹³, which acts as a sponge to remove matching scnRNAs from the active pool. Those that cannot find a match remain free to target homologous sequences in the zygotic MAC when it develops, thereby recruiting the IES excision machinery. This RNA-mediated genomic subtraction thus reproduces the deletions observed in the parental MAC and can account for epigenetic inheritance of alternative rearrangement patterns, such as retention of a given IES in the MAC^{14,15}, or deletion of a given gene¹⁶, across sexual generations.

In *Paramecium*, conjugation (the reciprocal fertilization of cells of opposite mating types) does not allow any significant exchange of cytoplasm between the mates, so that the pools of scnRNAs produced during meiosis are independently sorted in each cell. After fertilization, each developing MAC will thus reproduce the particular rearrangements present in the old MAC of its own cytoplasmic parent. This mechanism might underlie the maternal

(cytoplasmic) inheritance of mating types in *P. tetraurelia*, although this hypothesis could not be tested so far because the genes involved were unknown.

Although mating types were discovered in 1937¹⁷, to date the only available test relies on the ability of type O (Odd) to agglutinate with type E (Even) when vegetative cells become sexually reactive, a physiological state induced by mild starvation. Agglutination is a prerequisite for conjugation and occurs upon contact through adhesion of ciliary membranes. Mutational analyses showed that several genes are specifically required for expression of type E¹⁸⁻²⁰; mutations in these genes result in a constitutive O phenotype, which thus appears to be a default state. In *P. tetraurelia*, mating types are not genetically determined in the MIC²¹. Each new MAC becomes determined for one type during its development, and remains the same throughout vegetative growth of the derived clone. The O/E alternative is not random, but maternally inherited; elegant experiments showed that mating-type determination in the developing zygotic MAC is controlled through the cytoplasm²² by the maternal MAC^{23,24}.

We have identified a gene, *mtA*, which is specifically expressed in reactive cells of type E and encodes a ciliary transmembrane protein involved in agglutination. We show that mating types are determined during MAC development by an alternative rearrangement of *mtA*: in O clones, its promoter is excised as an IES in a scnRNA-dependent manner, preventing expression. We further found that in the sibling species *P. septaurelia*, a similar switch evolved in a different gene, *mtB*, which encodes a transcription factor required for *mtA* expression. These independent examples show that the scnRNA pathway, primarily a genome-defence mechanism, can be exapted to regulate cellular genes, and mediate epigenetic inheritance of essential phenotypic polymorphisms.

***mtA* rearrangements determine mating type**

To identify the putative mating type-determining gene, we used a whole-genome microarray to compare the transcriptomes of sexually reactive cells of both mating types (**Fig. 1a**). The gene found to have the largest E/O expression ratio, GSPATG00017533001, encodes a 1,275-aa protein with a signal peptide and cysteine-rich furin-like repeats followed by 5 transmembrane segments at the C terminus (**Fig. 1b** and **Supplementary Doc. 1a**); the N-terminal part is predicted to be outside the plasma membrane. This gene was later shown to be *mtA*¹⁹ (see below). Northern blot analyses confirmed that it is expressed only in E cells and further showed that expression is limited to sexual reactivity, since transcripts could not be detected in exponentially growing cells or in over-starved cells (**Fig. 1c**). Genes encoding structurally similar proteins were also found to be specifically expressed in sexually reactive cells, but in both mating types (**Fig. 1c**). RNAi-mediated silencing of *mtA* in E cells resulted in the default O phenotype during sexual reactivity (as determined by their capacity to agglutinate with E but not with O tester lines), indicating that *mtA* is required for E expression (**Supplementary Table 1**).

To understand the molecular basis for E-specific expression, we sequenced the *mtA* gene from the MACs of O and E cells, and the unrearranged MIC version. The MIC gene is interrupted by 4 IESs that are excised in both MAC types; however, a 195-bp segment containing the transcription start site and the first 26 bp of the coding sequence was found to be excised as an IES in O MACs, but retained in E MACs (**Fig. 2a-c** and **Supplementary Doc. 1a**). This segment contains the *mtA* promoter, as indicated by microinjection of different constructs into the MACs of O cells (**Extended Data Fig. 1a-c**): its presence upstream of the coding sequence was sufficient for transformed clones to express mating type E instead of mating type O.

An mtA-GFP fusion protein was detected in cilia of the anteroventral surface, although an excess of GFP fluorescence was seen in cytoplasmic structures, likely the ER (**Fig. 2d,e**). Cells expressing the fusion, unlike those expressing wild-type *mtA* transgenes, failed to agglutinate with O testers, suggesting that the GFP insertion (between furin-like repeats and transmembrane segments) interfered with the normal interactions and/or transport of the protein. However, they also failed to agglutinate with E testers, indicating that the GFP transgene was still able to suppress the endogenous mating-type O specificity. The fusion protein could be detected on cilia with anti-GFP antibodies in fixed, non-permeabilized cells (**Fig. 2f**), confirming that its N-terminal part is exposed outside of the ciliary membrane.

When *mtA*-transformed clones were taken through autogamy (a self-fertilization sexual process), the induced type E was robustly transmitted to progeny (**Extended Data Fig. 1b**). As observed for maternally controlled IESs^{14,15}, a plasmid containing only the 195-bp segment was sufficient to inhibit excision of the homologous sequence during development of zygotic MACs, causing the progeny of transformed O cells to switch to E (**Extended Data Fig. 1d**). Retention of the *mtA* promoter thus recapitulates both aspects of developmental determination for E, namely the capacity to express mating type E during the vegetative phase and the capacity, after sexual events, to direct determination of new MACs for the same type.

Exceptions to maternal inheritance

We further tested the correlation between mating types and *mtA* promoter status in situations where cells are known to escape maternal inheritance. After autogamy of wild-type cells, spontaneous mating-type reversion occurs at a frequency of ~1/3,000 in the E-to-O direction, and ~1/50,000 in the O-to-E direction²⁵. Autogamy was induced in a mass culture of E cells, and the progeny was screened for conjugating pairs upon starvation to identify O revertants. One pair was isolated; after conjugation, each partner develops two new MACs which segregate to daughter cells during the first division. Of the 4 clones produced, two were O and had excised the *mtA* promoter (**not shown**), suggesting the initial O revertant had arisen through spontaneous excision in one E cell during autogamy.

O-to-E reversion is rare in wild-type cells, but can be induced by the Mendelian mutation *mtF^E*, which affects MAC development and makes determination for E constitutive in mutant homozygotes, regardless of the maternal mating type^{26,27} (**Extended Data Fig. 2**). This pleiotropic and temperature-sensitive mutation also causes a number of other phenotypes and was shown to impair the excision of at least one IES²⁸, suggesting each phenotype is due to a rearrangement defect. We tested mating types and *mtA* promoter excision in the sexual progeny of an *mtF^E/mtF⁺* clone of mating type O, with MAC development at permissive or restrictive temperatures. Autogamy of the heterozygote produces entirely homozygous F2 clones with a 50% probability to inherit either allele. In all cases, *mtF^E* homozygotes that switched to E showed at least partial retention of the *mtA* promoter, while O clones (wild-type clones, or mutant clones at the permissive temperature) showed complete excision (**Extended Data Fig. 2e**). We conclude that the *mtF^E* mutation induces E determination by affecting excision of the *mtA* promoter during MAC development.

Expression and inheritance uncoupled

Previous mutational analyses showed that the expression of mating types during sexual reactivity can be uncoupled from the mechanism that ensures their transgenerational inheritance^{19-21,23}. The recessive mutations *mtA^O*, *mtB^O*, and *mtC^O* all preclude expression of type E and restrict cells to the default O phenotype, but they do not affect the maternal

inheritance of mating-type determination: mutant homozygotes formed in an E cytoplasmic lineage, despite expressing O, will keep the memory of E determination for an indefinite number of sexual generations, as shown by the fact that their progeny switches back to E expression when the wild-type allele is reintroduced by conjugation¹⁹ (**Extended Data Fig. 3a,b**).

To determine whether this memory is attributable to retention of the *mtA* promoter, we crossed each of the mutants to the wild type. Autogamy of F1 heterozygotes of mating type E yielded homozygous F2 progeny with a 1:1 ratio of E- and O-expressing clones, reflecting the Mendelian segregation of wild-type and mutant alleles. Both types of clones retained the *mtA* promoter (**Extended Data Fig. 3c**). That the O-expressing mutant clones could still transmit E determination was verified by crossing them again to wild-type E cells: the second-round F1 heterozygotes derived from both parents always expressed E and retained the *mtA* promoter (**Extended Data Fig. 3d**). Autogamy of these F1s again resulted in the expected Mendelian segregation of E- and O-expressing clones among second-round F2s (**Supplementary Table 2**). Thus, independently of the mating type being expressed, maternal inheritance of *mtA* promoter retention may indeed underlie the epigenetic memory of E determination.

To understand how mutant clones can express mating type O despite retaining the *mtA* promoter, we tested *mtA* expression in E- and O-expressing lines, using F3 populations obtained by an additional autogamy of second-round F2s. In the case of the *mtA^O* mutation, northern blots showed that *mtA* (GSPATG00017533001) transcripts were produced in sexually reactive cultures of O-expressing lines (**Fig. 3a**), prompting us to resequence the gene in mutant clones. A substitution was found to change Arg codon 751 to a stop codon, and the mutation cosegregated with mating type O expression among second-round F2s (**Supplementary Table 2a**). This revealed that GSPATG00017533001 is *mtA* and confirmed that the encoded protein is required for expression of mating type E. The mutation lies at a distance from the *mtA* promoter and apparently does not affect the regulation of its excision, explaining how O-expressing mutant homozygotes can transmit either O or E determination.

In contrast to *mtA^O* lines, *mtB^O* and *mtC^O* O-expressing lines did not produce *mtA* mRNA upon sexual reactivity (**Fig. 3a**). The mutations were identified by whole-genome sequencing, and found to be substitutions in GSPATG00026812001 and GSPATG00009074001, respectively (**Supplementary Doc. 1b,c**). Correct identification of the mutations was confirmed by their co-segregation with O expression among F2 lines (**Supplementary Table 2c,d**), and by transformation of the MACs of E-determined, O-expressing mutants with the wild-type alleles, which resulted in *mtA* transcription and E expression in both cases (**Fig. 3b** and **Extended Data Fig. 4**). Thus, both *mtB* and *mtC* appear to be transcription factors required for *mtA* expression in E-determined cells. The *mtB* and *mtC* genes are constitutively expressed at low levels, and both of the encoded proteins are predicted to be nuclear. *mtC* is a 138-aa protein containing a C2H2 zinc finger, the structure of which is likely affected by the *mtC^O* mutation (**Supplementary Doc. 2**). Although the 310-aa *mtB* protein does not contain any recognizable domain, an *mtB*-GFP fusion protein was found to localize to the MAC (**Fig. 3c**). Deep sequencing of the transcriptomes of *mtB^O* and *mtC^O* mutants further showed that *mtA* is the only gene that requires both factors for expression during sexual reactivity; importantly, it is the only gene found to be down-regulated in the *mtB^O* mutant (**Supplementary Table 3**).

***mtA* promoter excision is regulated by scnRNAs**

Excision of the *mtA* promoter in O cells occurs between two 5'-TA-3' dinucleotides within a reasonable IES end consensus (**Supplementary Doc. 1a**), suggesting the general IES excision machinery is involved. RNAi-mediated depletion of the Pgm endonuclease⁵ during autogamy of O cells indeed impaired excision and resulted in accumulation of unexcised copies in the developing new MAC (**Fig. 4a**). So did depletion of proteins known to be involved in scnRNA biogenesis or action, including the Dicer-like Dcl2 and Dcl3¹², the Piwi-like Ptiwi01 and Ptiwi09¹¹, and the Nowa1 and Nowa2 RNA-binding proteins²⁹, indicating that the scnRNA pathway is required to target excision in O cells. Depletion of each pair of proteins causes massive retention of IESs genome-wide, resulting in non-functional new MACs and post-autogamous lethality. In conditions of partial depletion, obtained by silencing only one of the two Piwi (or the two Dicer-like) genes, the progeny is usually viable. Analysis of individual clones showed that they frequently retained the *mtA* promoter in the new MAC, and in all cases this correlated with a switch to mating type E (**Fig. 4b**).

No other IES is known to be affected in these conditions; in this regard, excision of the *mtA* promoter behaves like MAC deletions of cellular genes, which were shown to be more sensitive to partial impairment of the scnRNA pathway^{11,12}. While other IESs are intervening sequences that must be excised to reconstitute functional genes, the 195-bp segment must be retained in the MAC to produce a functional *mtA* gene. The homologous segment is not excised in *mtA* orthologs from other *P. aurelia* species (see below) or in distant paralogs such as *mtAL1* and *mtAL2*, suggesting that excision of a *bona fide* part of the MAC genome evolved as a derived character in *P. tetraurelia*, as mutations happened to create suitable Pgm cleavage sites allowing recruitment of the IES excision machinery.

Deep sequencing of small RNAs from an early conjugation time point confirmed that scnRNAs are produced from both strands of the *mtA* promoter, as they are from the rest of the germline genome (**Extended Data Fig. 5**). On average, scnRNA coverage was ~2-fold higher for IESs than for other genome features, as independently found in another study³⁰. Although we cannot exclude that this results from precocious degradation of MAC-matching scnRNAs, an intrinsically more abundant production of scnRNAs from IESs could explain why IES excision is less sensitive to partial impairment of the scnRNA pathway than gene deletion. Northern blot analyses of autogamy time courses showed that *mtA*-promoter scnRNAs are produced in similar amounts during meiosis of O and E cells and failed to detect any difference in the timing of their disappearance (**Extended Data Fig. 6**). Thus *mtA*-promoter scnRNAs may be inactivated in E cells by some other mechanism than the active degradation proposed in *T. thermophila*³¹⁻³³, for instance by sequestration in the maternal MAC. Whatever the precise mechanism, one prediction of the model is that RNAi-mediated destruction of maternal transcripts of the *mtA* promoter during autogamy of E cells will prevent inactivation of homologous scnRNAs, licensing them to target excision in the zygotic MAC¹³. Indeed, feeding E cells with double-stranded RNA homologous to the *mtA* promoter prior to autogamy induced its precise excision in the new MAC and resulted in O progeny (**Supplementary Table 4**). The induced type O was thereafter inherited for at least 3 sexual generations (**not shown**).

A different switch in *P. septaurelia*

P. tetraurelia belongs to a group of 15 sibling species that are morphologically indistinguishable but sexually incompatible^{1,34,35}. All have homologous O and E mating types, as shown by cross-agglutination between the O types of some species and the E types of others. However, maternal inheritance of mating types is observed in only half of these

species. In others, mating types are randomly determined during MAC development, without any influence of the maternal MAC, and in a single species they are genetically determined in the MIC by Mendelian alleles. The distribution of these systems in the phylogenetic tree of *aurelia* species suggests multiple changes during evolution of the complex^{36,37} (**Extended Data Fig. 7**). We first sequenced *mtA*, *mtB*, and *mtC* orthologs in two strains of *P. octaurelia*, a maternal-inheritance species closely related to *P. tetraurelia*, with which it can form viable (though sterile) F1 hybrids³⁸. We found the *mtA* promoter to be excised in O clones, with the same boundaries as in *P. tetraurelia*, but not in E clones (**Extended Data Fig. 8 and Supplementary Doc. 1a**). This excision system thus likely evolved in the common ancestor of the two species.

We then examined *P. septaurelia*, another maternal-inheritance species which groups with random-determination species. Sequencing of the *mtA*, *mtB*, and *mtC* orthologs in the wild-type strain 227 showed that the *mtA*²²⁷ promoter was not excised in the MACs of O cells (**Supplementary Doc. 1a**). Instead, MAC copies of the *mtB*²²⁷ gene contained either of two alternative deletions of coding-sequence segments between IES-like boundaries (**Fig. 5a and Supplementary Doc. 1b**), suggesting the gene was rearranged into non-functional forms. Our stock of strain 227 contained only O cells, but complementation of their MACs by microinjection of the *P. tetraurelia* *pmtB*⁵¹ transgene made these cells phenotypically E and resulted in *mtA* transcription during sexual reactivity (**Fig. 5b**). Thus, mating type E is characterized by *mtB*-dependent *mtA* expression in *P. septaurelia*, as in *P. tetraurelia*.

The *mtB* gene belongs to a family of up to 6 paralogs in *P. aurelia* species, and sequence conservation indicates that all have been under purifying selection since before speciation (**not shown**). The excision events that inactivate *mtB*²²⁷ in O cells are not observed in any other paralog or in *mtB* orthologs from other species, where the 5' TAs correspond to different nucleotides. As with *mtA* promoter excision in *P. tetraurelia*, this excision system apparently does not originate from any TE insertion but seems to have arisen in *P. septaurelia* following the chance appearance of good matches to the IES end consensus, which appears to be conserved among *P. aurelia* species³⁹.

Taking advantage of the evidence that the *mtB*⁵¹ protein from *P. tetraurelia* can activate the *mtA*²²⁷ promoter in *P. septaurelia*, we microinjected the *mtA*²²⁷ gene into the MACs of *P. tetraurelia* cells of mating type O, and tested the sexual preference of transformed clones with O and E testers of both species (**Fig. 5c**). Consistent with the reported lack of cross-agglutination between the complementary types of these two species¹, uninjected control clones reacted only with *P. tetraurelia* E testers. In contrast, transformed clones agglutinated only with *P. septaurelia* O testers, allowing interspecific conjugation to occur. Thus *mtA*²²⁷ is sufficient for expression of the *P. septaurelia* E specificity, incidentally confirming that *mtA* localizes on the external side of ciliary membranes. The lack of reaction of transformed clones with *P. tetraurelia* O testers indicates that *mtA* interacts with an unknown O-specific receptor which also differs between species. Furthermore, transformed clones did not react with *P. tetraurelia* E testers, implying that *mtA*²²⁷ expression, like *mtA*⁵¹ expression (**Extended Data Fig. 1**), functionally masks expression of the endogenous O-specific receptor. This would explain the default nature of the O type.

Strain 38 is a natural *P. septaurelia* isolate carrying a Mendelian mutation at a single locus (*mt*) which, in contrast to the *P. tetraurelia* mutations, affects both the expression and the epigenetic inheritance of mating types²⁰. This dual effect suggests that *mt* is the gene that controls mating types through an alternative rearrangement in that species (see **Extended Data Fig. 9**), providing an independent means to confirm the identification of the *P. septaurelia* switch. Sequencing of the *mtB*³⁸ allele indeed revealed mutations that can explain both effects (**Extended Data Fig. 9 and Supplementary Doc. 1b**), and the analysis of a cross

between strains 38 and 227 showed that *mt* is identical or closely linked to *mtB* (**Supplementary Table 5**). To directly confirm that *mtB* rearrangements are regulated by homology-dependent maternal effects, the full-length, functional MAC form of the *mtB*²²⁷ allele was microinjected into the MACs of O cells of strain 227. Transformed clones expressed E and produced E post-autogamous progeny in 3 of 4 cases (**Extended Data Fig. 10**). We conclude that mating types are determined in *P. septaurelia* by maternally inherited alternative rearrangements of the *mtB* gene, a mechanism which must have evolved independently from that of *P. tetraurelia*.

Discussion

Seventy-five years after mating types were first described in *P. tetraurelia*, the expression of a single gene, *mtA*, was found to make the difference between the two types. *mtA* transcription in E cells was further shown to require the *mtB* and *mtC* gene products, and this E-specific pathway appears to be conserved in other *P. aurelia* species. Early genetic analyses^{18,19} suggested that the O-specific ciliary ‘substance’ mediating adhesion to E cilia may be the precursor of the E-specific ‘substance’, a metabolic conversion which would require the *mtA*, *mtB*, and *mtC* gene products. However, none of the 3 proteins has recognizable catalytic activity, and *mtA* is a ciliary transmembrane protein directly involved in species-specific recognition of O cells. While the nature of the O-specific receptor remains unknown, its expression or function appears to be masked by the *mtA* protein, given the default O phenotype observed in mutants. Expressing the *P. septaurelia* *mtA* ortholog in O cells of *P. tetraurelia* indeed not only provided them with the *P. septaurelia* E specificity, but also blocked expression of the *P. tetraurelia* O specificity. In the related *P. caudatum*, an antigen involved in agglutination was shown to be produced in the cytoplasm of both O and E reactive cells, but to localize in the ciliary membrane only in O cells⁴⁰. In E cells of *P. aurelia* species, the *mtA* protein could inhibit the synthesis, processing or transport of the O-specific receptor. *mtA* is structurally similar to the proteins recently shown to be specific for each of the 7 mating types of *T. thermophila*⁴¹, but how these function in mating-type recognition remains to be determined.

Despite conservation of the *mtA* on/off switch as the E/O expression determinant, the mechanisms used to produce O clones differ among *P. aurelia* species. Excision of the *mtA* promoter is shared between the closely related *P. tetraurelia* and *P. octaurelia*, but in *P. septaurelia* *mtA* expression is prevented by deletions in *mtB* (**Fig. 6**). This appears to be functionally equivalent since *mtB* is highly specific for the *mtA* promoter in *P. tetraurelia*. The *P. tetraurelia* rearrangement involves the IES excision machinery and is regulated by the scnRNA pathway, explaining the maternal inheritance of mating types in this species, and probably in the other two as well. Ciliate scnRNAs resemble metazoan piRNAs in many respects, including their central roles in epigenetic reprogramming events that are critical for sexual reproduction⁴², and both have been described as genomic ‘immune systems’ primarily involved in the control of TEs^{2,43-45}. One difference, however, is that metazoan piRNA systems keep a memory of ‘non-self’ sequences, based on integration into specific small RNA-producing loci^{43,45,46}. The *Paramecium* system appears to work the opposite way, keeping a memory of ‘self’ sequences in the form of the maternal MAC, which is probed by genome-wide scnRNAs at each generation to identify ‘non-self’ molecules by subtraction.

As a consequence, the *Paramecium* genome-defence mechanism can easily be recruited for heritable regulation of *bona fide* cellular genes, an important question that has been raised in diverse eukaryotes^{45,47-50}. Indeed, the sequences being excised from the *mtA* and *mtB* genes differ from other IESs^{3,7} in that they are functional parts of cellular genes and apparently do

not derive from the insertion of TEs; moreover, the *mtA* rearrangement behaves like deletions of cellular genes in its mechanistic requirements, which show subtle differences with IES excision^{11,12}. The independent evolution of the *P. tetraurelia* and *P. septaurelia* switches suggests that exaptation of the scnRNA pathway is a general mechanism for transgenerational epigenetic inheritance of differentiated states. In plants and mammals, epigenetic regulation of cellular genes often relies on transposon-induced epigenetic changes^{47,48}, as is the case of sex determination in melon⁵¹. The capacity of the *Paramecium* mechanism to inactivate any single-copy gene makes it even more flexible, dispensing with the need for transposon insertion. In keeping with the low information content of the IES end consensus, accidental rearrangement errors are not rare genome-wide^{3,39,52}, and may occasionally provide a selective advantage in specific conditions; maternal transmission to sexual progeny would thus allow continuous adaptation of the somatic (MAC) genome, independently of any Mendelian (MIC) mutation.

Methods Summary

Unless otherwise stated, all experiments were carried out with the entirely homozygous strain 51 of *P. tetraurelia*; the origins of other strains and species are given in the online Methods section. Molecular biology experiments used standard procedures. *Paramecium*-specific methods (RNAi-mediated gene silencing by dsRNA feeding, DNA microinjection into the MAC) have been published and are referenced in the online Methods section. For mating-type tests, testers were prepared from cell lines of known mating types by refeeding ~1000 autogamous cells in tubes with 4 ml of 0.2 X WGP medium bacterized with *K. pneumoniae* (light medium) and incubating overnight at 27°C. The next day, tubes were refed with 8 ml of light medium and again incubated overnight at 27°C. The following day, reactive cells concentrated near the top of the tube were collected (~1.5 ml per tube), checked by mixing aliquots with reactive cells of the complementary type, and used in mating type tests. Mass post-autogamous progenies to be tested were made reactive in the same way. To test individual clones, single karyonides (out of autogamy or conjugation) were isolated in 250 µl of light medium and incubated until starved. They were then refed with 250 µl and tested the next day. Bioinformatic analyses and statistical testing (microarray expression data, whole-genome transcriptome profiling by RNAseq) are described in the online Methods section.

References

1. Sonneborn, T.M. *Paramecium aurelia*. in *Handbook of Genetics*, Vol. Plenum Press (ed. King, R.C.) 469-594 (New York, 1974).
2. Chalker, D.L. & Yao, M.C. DNA elimination in ciliates: transposon domestication and genome surveillance. *Annu Rev Genet* **45**, 227-46 (2011).
3. Arnaiz, O. *et al.* The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet* **8**, e1002984 (2012).
4. Betermier, M. Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate *Paramecium*. *Res Microbiol* **155**, 399-408 (2004).
5. Baudry, C. *et al.* PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev* **23**, 2478-83 (2009).

6. Klobutcher, L.A. & Herrick, G. Consensus inverted terminal repeat sequence of Paramecium IESs: resemblance to termini of Tc1-related and Euplotes Tec transposons. *Nucleic Acids Res* **23**, 2006-13 (1995).
7. Dubois, E. *et al.* Transposon Invasion of the Paramecium Germline Genome Countered by a Domesticated PiggyBac Transposase and the NHEJ Pathway. *Int J Evol Biol* **2012**, 436196 (2012).
8. Coyne, R.S., Lhuillier-Akakpo, M. & Duharcourt, S. RNA-guided DNA rearrangements in ciliates: is the best genome defence a good offence? *Biol Cell* **104**, 309-25 (2012).
9. Duharcourt, S., Lepere, G. & Meyer, E. Developmental genome rearrangements in ciliates: a natural genomic subtraction mediated by non-coding transcripts. *Trends Genet* **25**, 344-50 (2009).
10. Meyer, E. & Chalker, D.L. Epigenetics of ciliates. in *Epigenetics* (eds. Allis, C.D., Jenuwein, T., Reinberg, D. & Caparros, M.C.) 127-150 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2007).
11. Bouhouche, K., Gout, J.F., Kapusta, A., Betermier, M. & Meyer, E. Functional specialization of Piwi proteins in Paramecium tetraurelia from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res* **39**, 4249-64 (2011).
12. Lepere, G. *et al.* Silencing-associated and meiosis-specific small RNA pathways in Paramecium tetraurelia. *Nucleic Acids Res* **37**, 903-15 (2009).
13. Lepere, G., Betermier, M., Meyer, E. & Duharcourt, S. Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in Paramecium tetraurelia. *Genes Dev* **22**, 1501-12 (2008).
14. Duharcourt, S., Butler, A. & Meyer, E. Epigenetic self-regulation of developmental excision of an internal eliminated sequence on Paramecium tetraurelia. *Genes Dev* **9**, 2065-77 (1995).
15. Duharcourt, S., Keller, A.M. & Meyer, E. Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in Paramecium tetraurelia. *Mol Cell Biol* **18**, 7075-85 (1998).
16. Garnier, O., Serrano, V., Duharcourt, S. & Meyer, E. RNA-mediated programming of developmental genome rearrangements in Paramecium tetraurelia. *Mol Cell Biol* **24**, 7370-9 (2004).
17. Sonneborn, T.M. Sex, Sex Inheritance and Sex Determination in Paramecium Aurelia. *Proc Natl Acad Sci U S A* **23**, 378-385 (1937).
18. Butzel, H.M. Mating Type Mutations in Variety 1 of Paramecium Aurelia, and Their Bearing upon the Problem of Mating Type Determination. *Genetics* **40**, 321-30 (1955).
19. Byrne, B.C. Mutational Analysis of Mating Type Inheritance in Syngen 4 of PARAMECIUM AURELIA. *Genetics* **74**, 63-80 (1973).
20. Taub, S.R. The Genetic Control of Mating Type Differentiation in Paramecium. *Genetics* **48**, 815-34 (1963).
21. Sonneborn, T.M. Genetics of cellular differentiation: stable nuclear differentiation in eucaryotic unicells. *Annu Rev Genet* **11**, 349-67 (1977).
22. Sonneborn, T.M. Recent advances in the genetics of Paramecium and Euplotes. *Adv Genet* **1**, 263-358 (1947).
23. Nanney, D.L. Mating type inheritance at conjugation in variety 4 of Paramecium aurelia. *J. Protozool.* **4**, 89-95 (1957).
24. Sonneborn, T.M. Patterns of nucleo-cytoplasmic integration in Paramecium. *Caryologia* **6 (Suppl.)**, 307-325 (1954).
25. Brygoo, Y. Genetic Analysis of Mating-Type Differentiation in PARAMECIUM TETRAURELIA. *Genetics* **87**, 633-53 (1977).

26. Brygoo, Y. & Keller, A.M. Genetic analysis of mating type differentiation in *Paramecium tetraurelia*. III. A mutation restricted to mating type E and affecting the determination of mating type. *Dev. Genet.* **2**, 13-22 (1981).
27. Brygoo, Y. & Keller, A.M. A mutation with pleiotropic effects on macronuclearly differentiated functions in *Paramecium tetraurelia*. *Dev. Genet.* **2**, 23-34 (1981).
28. Meyer, E. & Keller, A.M. A Mendelian mutation affecting mating-type determination also affects developmental genomic rearrangements in *Paramecium tetraurelia*. *Genetics* **143**, 191-202 (1996).
29. Nowacki, M., Zagorski-Ostojka, W. & Meyer, E. Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in *Paramecium tetraurelia*. *Curr Biol* **15**, 1616-28 (2005).
30. Sandoval, P.Y., Swart, E.C., Arambasic, M. & Nowacki, M. Functional Diversification of Dicer-like Proteins and Small RNAs Required for Genome Sculpting. *Dev Cell*, in press, doi:<http://dx.doi.org/10.1016/j.devcel.2013.12.010> (2014).
31. Aronica, L. *et al.* Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in *Tetrahymena*. *Genes Dev* **22**, 2228-41 (2008).
32. Mochizuki, K. & Gorovsky, M.A. Conjugation-specific small RNAs in *Tetrahymena* have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev* **18**, 2068-73 (2004).
33. Schoeberl, U.E., Kurth, H.M., Noto, T. & Mochizuki, K. Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in *Tetrahymena*. *Genes Dev* **26**, 1729-42 (2012).
34. Catania, F., Wurmser, F., Potekhin, A.A., Przybos, E. & Lynch, M. Genetic diversity in the *Paramecium aurelia* species complex. *Mol Biol Evol* **26**, 421-31 (2009).
35. Coleman, A.W. *Paramecium aurelia* revisited. *J Eukaryot Microbiol* **52**, 68-77 (2005).
36. Hall, M.S. & Katz, L.A. On the nature of species: insights from *Paramecium* and other ciliates. *Genetica* **139**, 677-84 (2011).
37. Phadke, S.S. & Zufall, R.A. Rapid diversification of mating systems in ciliates. *Biol J Linn Soc* **98**, 187-197 (2009).
38. Haggard, B.W. Interspecies crosses in *Paramecium aurelia* (syngen 4 by syngen 8). *J Protozool* **21**, 152-9 (1974).
39. Catania, F., McGrath, C.L., Doak, T.G. & Lynch, M. Spliced DNA sequences in the *Paramecium* germline: their properties and evolutionary potential. *Genome Biol Evol* **5**, 1200-11 (2013).
40. Xu, X., Kumakura, M., Kaku, E. & Takahashi, M. Odd mating-type substances may work as precursor molecules of even mating-type substances in *Paramecium caudatum*. *J Eukaryot Microbiol* **48**, 683-9 (2001).
41. Cervantes, M.D. *et al.* Selecting one of several mating types through gene segment joining and deletion in *Tetrahymena thermophila*. *PLoS Biol* **11**, e1001518 (2013).
42. Bourc'his, D. & Voinnet, O. A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science* **330**, 617-22 (2010).
43. Malone, C.D. & Hannon, G.J. Small RNAs as guardians of the genome. *Cell* **136**, 656-68 (2009).
44. Schoeberl, U.E. & Mochizuki, K. Keeping the soma free of transposons: programmed DNA elimination in ciliates. *J Biol Chem* **286**, 37045-52 (2011).
45. Siomi, M.C., Sato, K., Pezic, D. & Aravin, A.A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**, 246-58 (2011).

46. Khurana, J.S. *et al.* Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* **147**, 1551-63 (2011).
47. Castel, S.E. & Martienssen, R.A. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet* **14**, 100-12 (2013).
48. Daxinger, L. & Whitelaw, E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet* **13**, 153-62 (2012).
49. Luteijn, M.J. & Ketting, R.F. PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nat Rev Genet* **14**, 523-34 (2013).
50. Rouget, C. *et al.* Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo. *Nature* **467**, 1128-32 (2010).
51. Martin, A. *et al.* A transposon-induced epigenetic change leads to sex determination in melon. *Nature* **461**, 1135-8 (2009).
52. Duret, L. *et al.* Analysis of sequence variability in the macronuclear DNA of *Paramecium tetraurelia*: a somatic view of the germline. *Genome Res* **18**, 585-96 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements

We thank S. Malinsky, C. Ciaudo and M.-A. Félix for a critical reading of the manuscript, and S. Marker and all other lab members for continuous support and discussions. This work was supported by the “Investissements d’Avenir” program ANR-10-LABX-54 MEMO LIFE/ANR-11-IDEX-0001-02 Paris Sciences et Lettres* Research University and by grants ANR-08-BLAN-0233 “ParaDice” and ANR-12-BSV6-0017 “INFERNO” to EM/LS/SD, an “Equipe FRM” grant to EM, grants ANR-2010-BLAN-1603 “GENOMAC” and CNRS ATIP-Avenir to SD, and National Science Foundation grant MCB-1050161 to M. Lynch (JFG). DPS was supported by Ph.D. fellowships from the Erasmus Mundus program and from the Ligue Nationale Contre le Cancer. MLA was supported by Ph.D. fellowships from the Ministère de l’Enseignement Supérieur et de la Recherche and from the Fondation de la Recherche Médicale. AP was supported by grant RFBR 13-04-01683a. Some strains used in this study are maintained at the Centre of Core Facilities “Culture Collection of Microorganisms” in St Petersburg State University. The sequencing of the *mtB^O* and *mtC^O* MAC genomes benefited from the facilities and expertise of the high throughput sequencing platform of IMAGIF (Centre de Recherche de Gif - www.imagif.cnrs.fr). The *mtB^O* and *mtC^O* transcriptomes were sequenced at the IBENS genomics facility (Génomique Montagne Sainte-Geneviève). This study was carried out in the context of the CNRS-supported European Research Group “Paramecium Genome Dynamics and Evolution” and the European COST Action BM1102.

Author Contributions

DPS did most of the experimental work presented here and contributed to the design of experiments. BS characterized mRNAs and contributed to silencing experiments and northern blot analyses. GG contributed to gene sequencing, plasmid construction, PCR analyses and cell line maintenance. JFG did the microarray analysis, and AAF the confocal analysis of mtA-GFP fusions. AA, KL and JMA carried out the deep sequencing of small RNAs, and CB that of the *mtB^O* and *mtC^O* transcriptomes; OA and LS did the bioinformatic analyses. KB,

MLA, VT and SD showed the role of scnRNA pathway genes in *mtA* promoter excision. SB did the *mtA* promoter dsRNA feeding experiment. AP contributed to the analysis of the *mtA^O* mutant and provided *P. octaurelia* and *septaurelia* strains. MP contributed to the analysis of the *mtB^O* mutant and prepared samples from the *P. octaurelia* cross, which was carried out by EP. EM conceived the study and wrote the paper.

Author Information

Microarray data have been deposited at the Gene Expression Omnibus database⁵³ under accession number GSE43436 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=irmtmgmorpiblwj&acc=GSE43436>). RNASeq data (transcriptomes of *mtB^O* and *mtC^O* mutants) have been deposited in the European Nucleotide Archive (EBI) (<http://www.ebi.ac.uk/ena/data/view/ERP002291>). Small RNA sequences have been deposited at the EBI (<http://www.ebi.ac.uk/ena/data/view/ERP001812>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to emeyer@biologie.ens.fr.

Figure Legends

Figure 1. *mtA* and related genes. **(a)** Correlation of gene expression levels in sexually reactive cells of mating types O and E, as determined by NimbleGen microarray hybridization. The outlier indicated by an arrow is *mtA* (GSPATG00017533001). **(b)** Structural similarity of *mtA* and the related proteins *mtAL1* and *mtAL2*. **(c)** Northern blot analysis of expression of *mtA*, *mtAL1* (GSPATG00025159001) and *mtAL2* (GSPATG00002922001). The panels on the left compare total RNA samples from sexually reactive cultures of the two mating types (React., O and E). In the panels on the right, RNA samples were extracted from: V, vegetative, exponentially growing cells; R, mildly starved and sexually reactive; S, starved for 48 hrs and no longer reactive. The 3 mRNAs migrate just above the large rRNA, as can be seen from the pattern of background hybridization.

Figure 2. Structure of *mtA* in the MIC and in the MACs of E and O cells. **(a)** Black boxes are IESs; the 195-bp segment (grey box) is excised only in O MACs. A small unannotated gene downstream of *mtA* is conserved in other species. **(b)** Southern blot of *EcoRI*-digested total DNA, hybridized with an *mtA* probe (PCR5) revealing fragments of 1,945 and 2,140 bp in O and E clones. **(c)** PCR amplification of the *mtA* 5' end (PCR1) on independent pools of O and E clones (266 and 461 bp, respectively). pO and pE, control PCRs on cloned MAC versions; M, size markers. **(d-f)** Localization of an *mtA*-GFP fusion protein in an O cell transformed with a promoter-containing *mtA*-GFP transgene. **d:** DIC image. Scale bar, 20 μ m. **e:** Intracellular confocal optical section. Some GFP fluorescence (green) is detected in cilia of the anterior (top), but not posterior, part of the cell. The MAC is stained with Hoechst (blue). **f:** Confocal image of the ventral cell surface, showing the opening of the oral apparatus. The fixed, non-permeabilized cell was labeled with anti-GFP and secondary antibodies (red).

Figure 3. Molecular analysis of expression mutants. **(a)** *mtA* expression in E-determined *mtA^O*, *mtB^O*, and *mtC^O* homozygotes. The top panel shows a PCR analysis (PCR1) of E- and O-expressing F3 lines from the second round of backcross of each mutant to wild-type E cells. pO and pE, control PCRs on cloned MAC versions of *mtA*. All clones retained the *mtA* promoter. The wild-type (wt) or mutant (m) genotypes were confirmed by sequencing. The middle and bottom panels show northern blots of total RNA samples from sexually reactive cultures of each F3 line, hybridized with *mtA* or *mtAL1* probes. **(b)** Northern blot analysis of

mtA expression in sexually reactive cultures of E-determined *mtB^O* and *mtC^O* clones transformed with plasmids *pmtB⁵¹*-NGFP and *pmtC⁵¹*, respectively. Transformed clones expressed mating type E (see **Extended Data Fig. 4**, same clone numbers), while uninjected controls (C2 and C1) expressed mating type O. **(c)** Fluorescence microscopy of an *mtB^O* cell transformed with *pmtB⁵¹*-NGFP and expressing mating type E (GFP and DAPI filters). Scale bar in left panel, 20 μ m.

Figure 4. Genes required for excision of the *mtA* promoter in O cells. **(a)** PCR analysis (PCR5) of *mtA* promoter retention in mass progenies of O clones after RNAi-mediated silencing of the indicated genes. Total DNA samples were prepared from starved post-autogamous cells when the new MACs were clearly visible. Because the parental MAC is still present at this stage, the promoter-excised version is amplified in all cases; the promoter-retaining fragment can be detected only if it accumulates in the zygotic MACs. ND7, unrelated-gene negative control; D2, *DCL2*; D3, *DCL3*; C1, no-silencing negative control; C2, empty-vector RNAi control; P1, *PTIWI01*; P9, *PTIWI09*; CE, control PCR5 on E cells. **(b)** PCR analysis (PCR5) and mating-type tests (below) of individual viable post-autogamous clones from non-lethal silencing conditions (including partial silencing of *NOWA1*).

Figure 5. *mtB* rearrangements determine mating types in *P. septaurelia*. **(a)** MIC and MAC versions of *mtB²²⁷*. The MIC sequence contains two IESs (black boxes) at the same positions as in *P. tetraurelia* strain 51. In the MAC of O cells of strain 227, the *mtB²²⁷* coding sequence suffers two alternative deletions, each between TA dinucleotides within IES-end consensus sequences (TACAG or TATAG, see **Supplementary Doc. 1b**). **(b)** Northern blot analysis of *mtA* expression in a sexually reactive, *pmtB⁵¹*-transformed clone of strain 227 (T). C, control uninjected O clone of strain 227. Tested mating types are indicated below the lanes. **(c)** Transformation of *P. tetraurelia* clones of mating type O with the *P. septaurelia mtA²²⁷* gene. A duplex PCR was used to assess transgene copy numbers through the relative abundance of a 176-bp transgene-specific product (*pmtA²²⁷*, PCR12), compared to the 266-bp product from the promoterless endogenous *mtA* (*mtA⁵¹*, PCR1). In highly transformed clones, the latter is not detectable. C1 and C2, uninjected control clones. Each clone was tested with O and E testers from *P. tetraurelia* strain 51 and from *P. septaurelia* strain 227. Number of +s represents strength of mating reaction; -, no agglutination. Clone 4, with a very low transgene copy number, reacted with both 51E and 227O testers.

Figure 6. A general model for mating-type determination in *P. aurelia* species. In the 3 species examined, mating type E depends on expression of the *mtA* protein during sexual reactivity. *mtA* transcription in turn requires the *mtB* and *mtC* gene products (the requirement for *mtC* in *P. septaurelia*, and for both genes in *P. octaurelia*, remains to be verified). In *P. tetraurelia* and *P. octaurelia*, mating type O is determined during MAC development by excision of the *mtA* promoter as an IES, preventing expression of the gene. In *P. septaurelia*, mating type O is determined by the excision of segments of the *mtB* coding sequence as IESs, which similarly prevents *mtA* expression.

Methods

Paramecium strains and cultivation

Unless otherwise stated, all experiments were carried out with the entirely homozygous strain 51 of *P. tetraurelia*. The mutant strains *mtA^O*, *mtB^O*, *mtC^O*, and *mtF^E*, as well as strain 32 of *P. tetraurelia*, were from the stock collection of the Centre de Génétique Moléculaire in Gif-sur-Yvette, France. The *mtA^O* mutant was independently ordered from ATCC (ATCC® Number : 30762™). Strains 138 and GFg-1 of *P. octaurelia* were from the stock collection of the

Institute of Systematics and Evolution of Animals, Krakow, Poland. Strain 227 from *P. septaurelia* was from the stock collection of the Centre of Core Facilities “Cultivation of Microorganisms” in St Petersburg State University, Russia. Strain 38 of *P. septaurelia* was ordered from ATCC (ATCC[®] Number: 30575[™]). Cultivation and autogamy were carried out at 27°C (unless otherwise stated) as described^{54,55}.

DNA and RNA extraction, Southern and northern blots

DNA and RNA samples were typically extracted from 200-400-ml cultures of exponentially growing cells at <1000 cells/ml or of autogamous cells at 2000-4000 cells/ml as previously described⁵. Small-scale DNA samples were prepared from ≤1,000 cells using the NucleoSpin[®] Tissue kit (Macherey-Nagel). The TRIzol (Invitrogen) RNA extraction procedure was adapted for small-scale cultures (~20-ml) of individual transformed clones during sexual reactivity. Electrophoresis and blotting were carried out according to standard procedures. For small-RNA northern blots, small RNAs were enriched by PEG8000 precipitation, and the equivalent of 75 µg of total RNA for each time point was run on a 15% polyacrylamide-urea gel, transferred on Hybond[™] NX membranes, and chemically cross-linked⁵⁶.

Microarray expression data

Expression data were obtained from single-channel NimbleGen microarrays covering all 39,642 annotated genes, with six different 50-mer probes per gene. Raw signals were processed using the standard RMA method⁵⁷. This includes a first step of background subtraction for each array, followed by between-array normalization which was carried out using the normalizeBetweenArrays function from the limma package⁵⁸. The latter step adjusts signals so that expression values have similar distributions in the two arrays considered in the analysis. The expression level of each gene was taken as the median signal from the six probes. The microarray platform has been described in more detail elsewhere⁵⁹.

Alignment and prediction programs

DNA and protein sequences were aligned using the MUSCLE software on the phylogeny.fr web site⁶⁰. Prediction of protein localization and transmembrane protein topology used the PSORT II⁶¹ and PolyPhobius⁶² servers.

Mating-type tests

Testers were prepared from cell lines of known mating types by refeeding ~1000 autogamous cells in tubes with 4 ml of 0.2 X WGP medium bacterized with *K. pneumoniae* (light medium) and incubating overnight at 27°C. The next day, tubes were refed with 8 ml of light medium and again incubated overnight at 27°C. The following day, reactive cells concentrated near the top of the tube were collected (~1.5 ml per tube), checked by mixing aliquots with reactive cells of the complementary type, and used in mating type tests. Mass post-autogamous progenies to be tested were made reactive in the same way. To test individual clones, single karyonides (out of autogamy or conjugation) were isolated in 250 µl of light medium and incubated until starved. They were then refed with 250 µl and tested the next day.

RNAi-mediated gene silencing by dsRNA feeding and microinjections

Procedures for RNAi and DNA microinjection into the MAC, as well as inserts used to silence scnRNA-pathway genes and *PGM* and the procedure to silence genes during autogamy, have been described^{5,11,12,29,63,64}. To silence genes during sexual reactivity, cells

were made reactive as described under ‘Mating-type tests’, except that the second refeeding with 8 ml was done with dsRNA feeding medium instead of normal WGP medium.

Plasmid constructs and probes

The plasmids used in this study are described in the relevant figures or Extended Data. The *mtA*-GFP fusion transgene used in **Fig. 2d-f** is a modified version of *pmtA*-ES (**Extended Data Fig. 1**) in which the EGFP coding sequence, flanked by Gly-Ser-Gly-Gly and Gly-Gly linkers, was inserted in place of *mtA* residue Asn1002. The *pmtA*²²⁷ transgene used in **Fig. 5c** is the *P. septaurelia* equivalent of *pmtA*-ES, with complete flanking intergenic regions (without the downstream unannotated gene). Complete sequences of all plasmids are available upon request. The *mtA* probe used to hybridize northern blots covered the whole coding sequence, except on the left panel of **Fig. 1c** where it spanned positions 415-1010 of the corrected gene model (PTETG5300016001, see ParameciumDB). The *P. septaurelia* probe in **Fig. 5b** and **Extended Data Fig. 9d** covered region 78-2245 of the *mtA*³⁸ gene model. For other genes, probes covered the indicated regions of each gene model: *mtAL1*, 3053-3599 of GSPATG00025159001; *mtAL2* : 3328-3790 of GSPATG00002922001; *PTIWI09*, 50-439 of GSPATG00020796001; *NOWA2*, 2513-3330 of GSPATG00016668001; *PGM*, 1540-2390 of GSPATG00016627001; *PTIWI10*, 4-264 of GSPATG00009468001. Oligonucleotide probes for scaffold-22 endogenous siRNAs (**Extended Data Fig. 6c**) have been described⁶⁵.

Confocal analysis of *mtA*-GFP localization

Cells were fixed in 2% paraformaldehyde in PBS buffer and rinsed in the same buffer supplemented with 3% BSA. All subsequent steps were performed in this buffer: cells were incubated with a commercial anti-GFP rabbit serum (Molecular Probes® A6455, Invitrogen) at a 1/1000 dilution, rinsed twice, incubated with an Alexa 568-coupled anti-rabbit antibody (Molecular Probes® goat anti-rabbit IgG(H+L) A11011, Invitrogen), rinsed in 1 mg/ml Hoechst 33258, mounted in Citifluor (Citifluor Ltd London), and observed under a Leica Confocal SP8 microscope with 405, 488 and 552 laser line excitations for blue, GFP and DIC, and red detection, respectively. Images were merged using ImageJ and Adobe Photoshop.

Deep sequencing of scnRNAs

Small-RNA libraries were previously constructed using 24-nt adaptors¹². The corresponding PCR products were extended with 6 degenerate nucleotides at each end and ligated to Illumina adaptors essentially following Illumina’s recommendations (TruSeq DNA sample prep kit protocol) before sequencing on the GAIIx system (100 nt single reads) (Illumina, USA).

Whole-genome transcriptome profiling

Eight Illumina sequencing libraries were constructed (2 for *mtB*^O, 2 for *mtC*^O, and 2 wild-type controls for each) using the Illumina TruSeq RNA Sample Preparation Kit. Approximately 30 million 101 nt read pairs were obtained for each library with the Illumina HiSeq 1000 sequencing system. Reads were processed to remove adapters and trimmed for quality before mapping to the strain 51 genome^{3,66} using BWA software⁶⁷. The fragments that mapped to each annotated gene⁶⁶ were counted only if both reads in the pair mapped to the same gene and only if the mapping for both reads was unique in the genome. Fragments were only counted if neither read in the pair had more than two mismatches. Fragment counts per gene were obtained from the BWA output using Samtools⁶⁸ and a custom Perl script. Surface antigen genes were excluded from the analysis since the cell cultures were not controlled for serotype and surface antigen mRNA can account for 3% of total cellular mRNA. Differential

gene expression was determined from fragment counts using the DESeq package⁶⁹. *P* values were corrected for multiple testing using the Benjamini-Hochberg procedure.

Additional References

53. Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-10 (2002).
54. Beisson, J. *et al.* Mass culture of *Paramecium tetraurelia*. *Cold Spring Harb Protoc* **2010**, pdb prot5362 (2010).
55. Beisson, J. *et al.* Maintaining clonal *Paramecium tetraurelia* cell lines of controlled age through daily reisolation. *Cold Spring Harb Protoc* **2010**, pdb prot5361 (2010).
56. Pall, G.S. & Hamilton, A.J. Improved northern blot method for enhanced detection of small RNA. *Nat Protoc* **3**, 1077-84 (2008).
57. Irizarry, R.A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-64 (2003).
58. Smyth, G.K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31**, 265-73 (2003).
59. Arnaiz, O. *et al.* Gene expression in a paleopolyploid: a transcriptome resource for the ciliate *Paramecium tetraurelia*. *BMC Genomics* **11**, 547 (2010).
60. Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**, W465-9 (2008).
61. Nakai, K. & Horton, P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* **24**, 34-6 (1999).
62. Kall, L., Krogh, A. & Sonnhammer, E.L. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics* **21** Suppl 1, i251-7 (2005).
63. Beisson, J. *et al.* DNA microinjection into the macronucleus of *paramecium*. *Cold Spring Harb Protoc* **2010**, pdb prot5364 (2010).
64. Beisson, J. *et al.* Silencing specific *Paramecium tetraurelia* genes by feeding double-stranded RNA. *Cold Spring Harb Protoc* **2010**, pdb prot5363 (2010).
65. Marker, S., Le Mouel, A., Meyer, E. & Simon, M. Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in *Paramecium tetraurelia*. *Nucleic Acids Res* **38**, 4092-107 (2010).
66. Aury, J.M. *et al.* Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**, 171-8 (2006).
67. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
68. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
69. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
70. Kapusta, A. *et al.* Highly precise and developmentally programmed genome assembly in *Paramecium* requires ligase IV-dependent end joining. *PLoS Genet* **7**, e1002049 (2011).

Extended Data Figure Legends

Extended Data Figure 1. The 195-bp segment contains the *mtA* promoter and ensures maternal inheritance of its own retention. **(a)** Plasmids used for microinjection into the MAC of O cells. The thick grey lines on either side of inserts represent vector sequences. Plasmid pmtA-E contains the E MAC version with entire intergenic regions; pmtA-ES lacks the downstream unannotated gene. pmtA-IES lacks intergenic sequences upstream of the 195-bp segment. pIES contains only the 195-bp segment. The PCRs used for testing transformation are shown. **(b)** Transformation of O cells with pmtA-E or pmtA-ES. PCR1 amplifies the promoter-containing form from either plasmid and the promoter-excised form from the endogenous *mtA* gene; relative amounts provide an indication of plasmid copy numbers. The mating types of injected clones and of their mass post-autogamous progeny are indicated. C1 and C2, uninjected controls; pE, control PCR on pmtA-E. **(c)** Transformation of O cells with pmtA-IES. The plasmid-specific PCR2 (1,533 bp) identifies transformed clones. The O type of clones 9 and 10 could be due to transgene silencing. **(d)** Transformation of O cells with pIES. Injected clones were tested with a duplex PCR : PCR4 amplifies a 450-bp product from the plasmid and PCR3 a 1,035-bp product from the endogenous *mtA* gene, hardly detectable in high-copy transformants. C1 and C2, uninjected controls; pD, control duplex PCR on an equimolar mix of pIES and pmtA-E. PCR1 revealed that the 195-bp segment was retained in the mass progeny of transformed clones; the selfer phenotype (S) of the mass progeny from clone 4 likely reflects heterogeneity among individual post-autogamous clones.

Extended Data Figure 2. Genetic analysis of mating-type inheritance in wild-type or *mtF^E* cells. **(a)** Mendelian segregation of a pair of alleles. In both conjugation and autogamy, the two MICs undergo meiosis, but a single haploid product is retained in each cell; an additional mitosis then produces two identical gametic nuclei. During conjugation of genetically different cells (m/m and +/+), reciprocal exchange of one gametic nucleus, followed by karyogamy, therefore results in genetically identical zygotic nuclei in the two exconjugants. The drawing shows the heterozygous MICs and MACs that develop from copies of the zygotic nuclei in each of the F1 cells. During autogamy, the two identical gametic nuclei fuse together, resulting in entirely homozygous zygotes; post-autogamous F2 progeny of heterozygotes have a 50% probability to keep each of the parental alleles. **(b)** Maternal (cytoplasmic) inheritance of mating types. Coloured Os and Es around the cells indicate the mating types expressed by vegetative clones in each sexual generation; the MACs are coloured to symbolize their mating-type determination states (O, blue; E, orange). Because very little cytoplasm is exchanged during conjugation, the parental MACs will condition zygotic MACs independently to be determined for the same mating types, resulting in cytoplasmic inheritance. **(c)** Cross of wild-type O cells to the determination mutant *mtF^E*. Mutant homozygotes are constitutively determined for E, even when the maternal MAC is determined for O. The black arrow in the MAC represents the correctly rearranged *G* gene; in mutant homozygotes, IES 51G4404 is retained in the MAC (yellow box). Note that the IES is also retained in wild-type F2 homozygotes deriving from the F1 heterozygote in the mutant cytoplasmic lineage, through a homology-dependent maternal effect. **(d)** Temperature sensitivity of the *mtF^E* mutation. Autogamy of an F1 heterozygote of mating type O (deriving from the wild-type parent), if performed at the permissive temperature of 14°C, yields F2 progeny that rarely show any phenotype. Most of the mutant homozygotes express mating type O even when tested at the restrictive temperature of 27°C, indicating that the *mtF^E* mutation does not prevent O expression. The pleiotropic phenotypes (E determination, retention of 51G4404) only become apparent after an additional round of autogamy, provided MAC development occurs at 27°C.

Extended Data Figure 3. Genetic analysis of mutations affecting mating-type expression. **(a)** Cross of O-determined expression mutants mtA^O , mtB^O , and mtC^O to wild-type E cells. F2 mutant homozygotes formed in the E cytoplasmic lineage express mating type O as a result of the mutation, but remain determined for E (orange MAC). **(b)** Cross of O-expressing, E-determined mutants mtA^O , mtB^O , and mtC^O (as produced in **a**) to wild-type E cells. E determination of the mutant parent is revealed by E determination and expression of the derived F1, which has received the wild-type allele. In this cross all cells are determined for E, and the expressed mating type simply depends on genotype. **(c)** mtA promoter retention and mating types of mtA^O/mtA^+ F1s from the first round of backcross of the mtA^O mutant with wild-type E cells, and of F2 progenies from F1 heterozygotes of mating type E. Top panel: PCR analysis (PCR5) of two sets of 4 F1 karyonides (from two pairs) showed that each set contained 2 O clones that had excised the mtA promoter, and likely derived from the mutant O parent, and two E clones that retained it, and likely derived from the wild-type E parent (mating types are indicated below each lane). Bottom panels: After autogamy of two E F1 heterozygotes (clone #4 in each set), 12 F2 homozygous progeny were isolated for each, grown and tested for mating-type expression and for mtA promoter retention using PCR6 (**Supplementary Table 6**), which amplifies products of 665 bp and 470 bp from the promoter-containing and promoter-excised versions, respectively). All clones retained the promoter, although a fraction of them (14/23) expressed mating type O, as expected for mtA^O homozygotes. ND, not determined. **(d)** mtA promoter retention and mating types of a typical set of 4 mtB^O/mtB^+ F1 karyonides from the second round of backcross (E-determined, O-expressing F2s from the first round of backcross were used as the mutant parents). PCR5 showed that the mtA promoter was now retained in F1 karyonides from both parents; all were of mating type E.

Extended Data Figure 4. Complementation of the mtB^O and mtC^O phenotypes with the wild-type alleles of GSPATG00026812001 and GSPATG00009074001, respectively. **(a)** Structure of the MIC and MAC versions of the mtB gene, and of the GFP fusion transgene used for complementation. The coding sequence (open arrow) is shown with the complete upstream and downstream intergenic regions. The MIC version contains two IESs (black boxes). Plasmid $pmtB^{51}$ -NGFP contains the MAC version with complete intergenic regions, and the EGFP coding sequence was fused at the 5' end of the mtB coding sequence. Thick grey lines on either side represent plasmid vector sequences. **(b)** PCR analysis and mating types of E-determined mtB^O mutant clones transformed with $pmtB^{51}$ -NGFP. PCR7 (top panel) amplifies products of 1148 bp from the plasmid, and of 419 bp from the endogenous mtB gene (**Supplementary Table 6**). The relative abundance of the two products gives an indication of plasmid copy number in each clone. C1 and C2, uninjected control clones. Mating types are indicated below each lane. Clones containing detectable plasmid amounts expressed mating type E, indicating that the GFP fusion protein is functional; the selfing phenotype (S) of clone 11 may be due to some cells having lost the plasmid. PCR1 (bottom panel) confirmed that all clones retained the mtA promoter. **(c)** Structure of the MIC and MAC versions of the mtC gene, and of the plasmid used for complementation. The coding sequence (open arrow) is shown with the complete upstream and downstream intergenic regions. The MIC version contains one IES (black box). Plasmid $pmtC^{51}$ contains the MAC version with 349 bp and 98 bp of upstream and downstream intergenic sequences, respectively. The plasmid-specific PCR8 amplifies a 419-bp product (**Supplementary Table 6**). **(d)** PCR analysis and mating types of E-determined mtC^O mutant clones transformed with $pmtC^{51}$. PCR8 (top panel) shows that all positive clones expressed mating type E. C1, C2 and C3, uninjected control clones. PCR1 (bottom panel) confirmed that all clones retained the mtA promoter.

Extended Data Figure 5. Deep sequencing of scnRNAs from an early conjugation time point (early meiosis). A total of 39,041,474 small-RNA sequences of 25 nt in length were obtained by Illumina sequencing of 4 libraries previously constructed from gel-purified molecules migrating at 23, 24, 25, or 26 nt (from the 7hND7 total RNA sample¹²). **(a)** 25-nt reads were first mapped (no mismatch allowed) on possible contaminant sequences (genomes of bacteria commonly found in cultures, *P. tetraurelia* mitochondrial genome, *P. tetraurelia* rDNA and other non-coding RNAs). The remainder was then mapped on known MIC sequences (the ‘MAC+IES’ genome, and 9 individual copies of the Sardine transposon³). To determine whether any of the remaining reads could correspond to IES excision junctions or to spliced transcripts, they were then mapped on the MAC genome and on the genome-wide set of spliced transcripts. Very few hits were found and these did not show the characteristic 5’-UNG signature of scnRNAs (see **d**), suggesting these molecules represent longer forms of endogenous siRNAs and/or could be mapped because of IES or intron annotation errors. Of the remaining unmapped reads (~49%), close to one half could be mapped on ‘PGM contigs’ a ~25-Mb preliminary assembly of MIC-specific sequences that are not collinear to MAC chromosomes³ and are thus likely to represent *bona fide* scnRNAs. **(b)** Statistics about the coverage of the ‘MAC+IES’ genome (17,786,284 reads). The average coverage is similar for exons, introns, and intergenic regions, but is ~2-fold higher for IES sequences. This may mean either that scnRNAs are initially produced in higher amounts from IESs, or that active degradation of scnRNAs homologous to MAC sequences is already under way at this early stage. **(c)** Mapping of 25-nt reads on the MIC version of the *mtA* gene region. Coding sequences of the corrected *mtA* gene model (PTETG5300016001, after correction of assembly indels and reannotation) and of the short gene downstream (PTETG5300018001) are shown as red boxes interrupted by introns and IESs. The yellow box represents the 195-bp segment of the *mtA* promoter that is excised in O cells; the 4 IESs are shown as green boxes. 25-nt reads mapping on the top or bottom strands of the region are colour-coded to indicate the number of times each one was sequenced : blue, one read; green, 2-9 reads; red, ≥10 reads. **(d)** Compositional profiles (nucleotide frequency on the left, and deviation from randomness on the right) of reads mapping to the MAC+IES genome, to the MAC genome, or to ‘PGM contigs’. For each set, logos are shown for all reads (left), or for the non-redundant subset only (right). For the ‘MAC+IES’ and ‘PGM contigs’ sets, the logos computed from all reads clearly show the 5’-UNG signature typical of scnRNA guide strands, as is the case for the major subset of reads starting with U (5’U), while the minor subset of sequences not starting with U (5’A/C/G), which may represent the steady state amount of passenger strands, shows the complementary signature CNA at positions 21-23. Deviation from randomness is greater when computed from all reads than when computed from the non-redundant subset only, suggesting that molecules with the signature are intrinsically produced in higher amounts, or are more stable. The small set of reads that mapped only to the MAC genome (putative IES excision junctions) does not show a clear 5’-UNG signature, suggesting that most of those are not scnRNAs. The same is true of the small number of reads mapping to exon-exon junctions (not shown).

Extended Data Figure 6. Northern blot analysis of *mtA*-promoter scnRNAs during autogamy of O or E cells. Mass cultures were allowed to starve, and RNA samples were extracted during exponential growth (Exp) and then at different times (0-20 hrs) after the appearance of the first meiotic cells. Cells become committed for autogamy at a fixed point of the cell cycle, so that the best synchrony that can theoretically be achieved is the duration of one cell cycle; in these experiments, the time between the first and the last cells to begin meiosis was ~12 hrs. **(a)** Proportions of cells in different cytological stages at each time point, as determined by

DAPI staining. Veg, vegetative cells; Mei, meiosis (crescent stage, meiosis I, meiosis II); Frg, cells with fragmented old MAC but new MACs not yet clearly visible; Dev, cells with two clearly visible developing new MACs; Kar, cells after the karyonidal division, with only one developing new MAC. **(b)** Northern blot analysis of mRNAs for early (*PTIWI09*, *NOWA2*), middle (*PGM*), or late (*PTIWI10*) genes. The same blots were hybridized successively with the 4 probes. **(c)** Northern blot analysis of small RNAs. The top panels show hybridization with the 195-bp *mtA*-promoter probe, revealing accumulation of 25-nt scnRNAs slightly later than expression of the meiosis-specific genes *PTIWI09* and *NOWA2*. As a control, the same blots were rehybridized with an oligonucleotide probe specific for a cluster of 23-nt endogenous siRNAs on scaffold 22^{12,65}, which are abundantly produced at all stages of the life cycle (bottom panels). Quantification of the *mtA*-promoter scnRNA signal and normalization with the siRNA signal did not reveal any significant difference in their amount or timing between the two mating types (not shown). Previous studies showed that the double-strand breaks that initiate IES excision in the new MACs start being detectable before the maximum of expression of the putative endonuclease *PGM*^{5,70} (no later than 10 hrs in these time courses). A PCR analysis of post-autogamous DNA samples confirmed that the *mtA* promoter was fully excised in mating type O, and fully maintained in mating type E (not shown).

Extended Data Figure 7. Phylogenetic tree of different strains from most *P. aurelia* species based on sequence polymorphisms in 3 nuclear genes (modified with permission from Fig. 2 of Catania *et al.* 2009, Mol. Biol. Evol. 26:421-431). Wrong species assignment of some strains studied here are corrected in red. Strain V5-13 was likely misassigned to *P. septaurelia* through the same error as for strain GFg-1 (see **Extended Data Fig. 8**). Species showing maternal inheritance or random determination of mating types are highlighted in red and blue, respectively.

Extended Data Figure 8. Strain GFg-1 belongs to the same species as strain 138, *i. e.* *P. octaurelia*. GFg-1 is among a set of strains that were originally assigned to *P. septaurelia* on the basis of conjugation tests with strain 38, a reference strain for that species. However, the stock of strain 38 used in these tests was in fact not 38, but instead some *P. octaurelia* strain, as shown by the comparison of *mtA* sequences with those from the original strain 38 obtained from ATCC (ATCC[®] Number: 30575TM) and those from strain 138, a reference strain for *P. octaurelia*. **(a)** Scheme of the cross GFg-1 (O) x 138 (E). The mating types of parents were determined by cross-agglutination with *P. tetraurelia* tester lines. The green and blue arrows besides each cell represent the *mtA* gene, colour-coded to indicate the GFg-1 and 138 alleles; the box at the 5' end symbolizes retention of the *mtA* promoter in the MAC genome on the E side of the cross. **(b)** Molecular characterization of two pairs of F1 heterozygotes, and of some F2 clones obtained by autogamy of these F1s. The parental origin of each F1 clone was ascertained by a sequence polymorphism in the mitochondrial COXII gene (PCR9, **Supplementary Table 6**). PCR amplification and sequencing of a segment of the *mtA* gene encompassing the promoter (PCR10, **Supplementary Table 6**) showed that the two F1 clones deriving from the GFg-1 parent were heterozygotes, and that the *mtA* promoter was precisely excised from both alleles (see **Supplementary Doc. 1a**). ND, not determined. Analysis of 6 viable F2 clones obtained by autogamy of the F1 clones deriving from the 138 parent showed that 4 of them were homozygous for the GFg-1 allele, while the other two were homozygous for the 138 allele; the *mtA* promoter was retained in all cases. The evidence for successful genetic exchange between strains GFg-1 and 138 and for viable recombinant F2 progeny demonstrates that these strains belong to the same species, *i. e.* *P. octaurelia*.

Extended Data Figure 9. Genetic and molecular analysis of the cross between strains 227 and 38 of *P. septaurelia*. **(a)** Maternal inheritance of mating types in the wild-type strain 227. **(b)** Strain 38 is genetically restricted to O expression, but constitutively determined for E. The mutation identified in strain 38 is of particular interest because it affects both the expression and the inheritance of mating types, which suggests that it lies in a gene that controls mating-type determination through an alternative rearrangement. Indeed, known *P. tetraurelia* mutations fall in two distinct categories. *mtA^O*, *mtB^O*, and *mtC^O* prevent expression of type E but have no effect on the rearrangement that determines mating types or on its maternal inheritance, while *mtF^E* lies in a *trans*-acting factor required for a subset of rearrangements during MAC development but has no effect on the expression of mating types during sexual reactivity. The only type of mutation that can be envisioned to affect both expression and determination/inheritance would be a mutation preventing expression of a functional protein required for E expression, and at the same time preventing in *cis* (by destroying a potential Pgm cleavage site) the rearrangement that normally inactivates this gene in the MAC of wild-type O cells. The *mt^{XIII}* allele of strain 38 restricts cells to O expression in a recessive manner, but also has a maternal effect that enforces constitutive E determination in sexual progeny. Interestingly, elegant experiments showed the latter effect to be dominant²⁰: the sexual progeny of a cell carrying at least one *mt^{XIII}* allele can never be determined for O or transmit O determination. **(c)** Sequencing of the *mtB³⁸* allele revealed features that may account for both effects. A frameshift mutation makes it a pseudogene, explaining the genetic restriction to O expression. In addition, a 6-bp deletion removes one of the IES-like boundaries used in the *mtB²²⁷* allele for coding-sequence deletions in O clones. Given the requirements of IES excision in *P. tetraurelia* and in sibling species³⁹, this can be predicted to make the same deletions impossible in the *mtB³⁸* allele, which would explain the constitutive E determination effect. The full-length *mtB³⁸* pseudogene in the MAC of 38 cells would indeed be expected, after a cross to 227, to protect the highly similar zygotic *mtB²²⁷* allele against coding-sequence deletions in the derived F1, through titration of homologous scnRNAs. **(d)** Molecular analysis of the 38 x 227 cross. To verify this maternal effect, we crossed an E-expressing 227 clone (*pmtB⁵¹*-transformed clone T, same as in Fig. 5b; C, uninjected control) with strain 38, and F1s were tested for mating types and for *mtA* expression by northern blotting. As expected, F1s deriving from the 38 parent (1b and 2b, as determined by sequencing of a mitochondrial polymorphism) were E and expressed *mtA*, implying that the incoming *mtB²²⁷* allele had been rearranged into a functional, full-length form in the MAC. After autogamy of F1 clone 2b, 24 independent F2 homozygotes were isolated and tested for mating types. Consistent with the Mendelian segregation of *mtB* alleles, 12 were O and 12 were E (**Supplementary Table 5**); northern blot analysis of 3 clones of each type showed that only E clones expressed *mtA*. **(e)** All F2 clones maintained the full-length *mtB* gene in the MAC. PCR11 (**Supplementary Table 6**) amplifies an 888-bp fragment from the MAC version of *mtB³⁸*, and an 893-bp fragment from the full-length MAC version of *mtB²²⁷*. C1 and C2, control PCR11 on two O clones of strain 227, showing the 806-bp and 744-bp fragments resulting from the two alternative coding-sequence deletions. **(f)** Mating types co-segregate with *mtB* alleles among F2s. Digestion of the PCR products with *AluI* distinguishes the 38 and 227 alleles. *mtA* and *mtC* alleles segregated independently (**Supplementary Table 5**).

Extended Data Figure 10. Transformation of *P. septaurelia* 227 O cells with the full-length *mtB²²⁷* gene induces heritable E expression. Plasmid *pmtB²²⁷*, containing the full-length E MAC form of *mtB²²⁷*, was microinjected into the MACs of O cells of strain 227, and 12 injected clones were tested by PCR11. This PCR amplifies an 893-bp product from the transgene, and products of 744 and 806 bp from the two internally deleted MAC forms of the endogenous *mtB²²⁷*. C1 and C2, control uninjected clones. Clones 4, 8, 9, and 11 contained

high copy numbers of the transgene, so that the PCR products from the endogenous gene are not detectable. Because only O tester lines were available, the mating types of injected clones could be ascertained only for E-expressing clones, as indicated below each lane. This was the case of the 4 high-copy clones; other clones were presumably O, although it was impossible to make sure that the cells were sexually reactive (indicated by '-'). Injected clones were then taken through autogamy, and the mass progenies were again tested for mating types. In this case, the mass progenies from clones 9 and 11, which proved to be pure E, were used as E testers to carry out a full test for some of the other progenies. Of the other two transformed clones expressing E, clone 4 gave rise to a selfing progeny (S), likely a mix of O and E clones, while clone 8 gave rise to a pure O progeny.

Figure 1

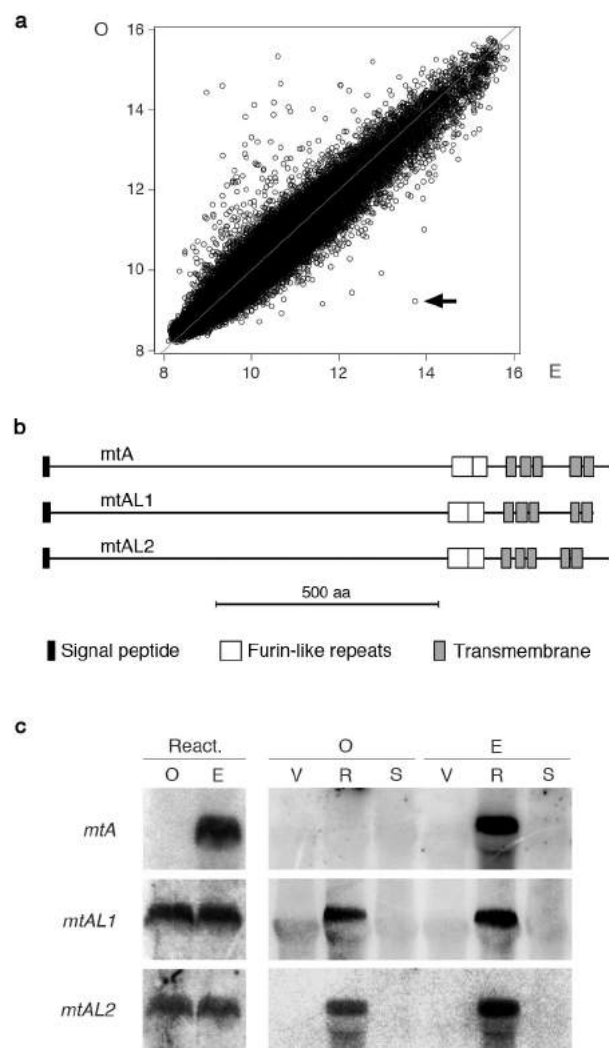


Figure 2

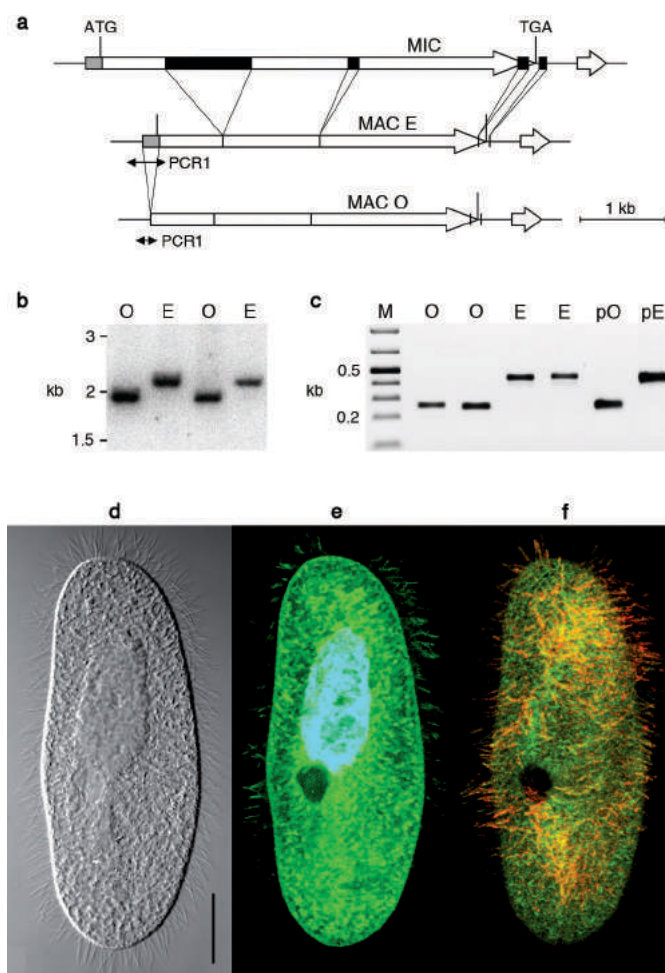


Figure 3

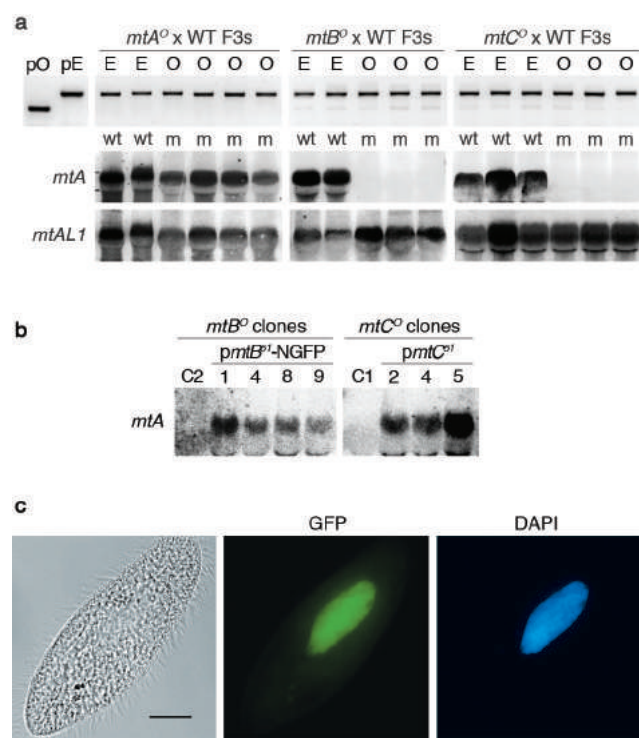


Figure 4

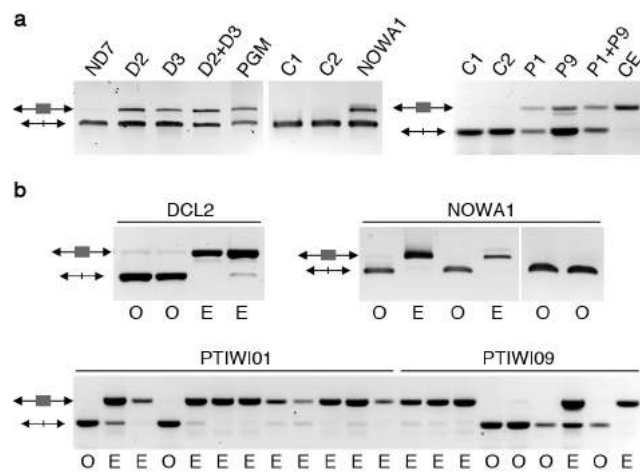


Figure 5

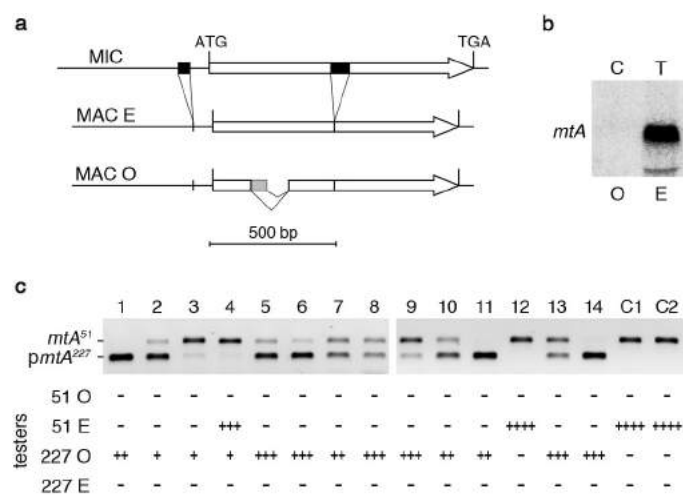
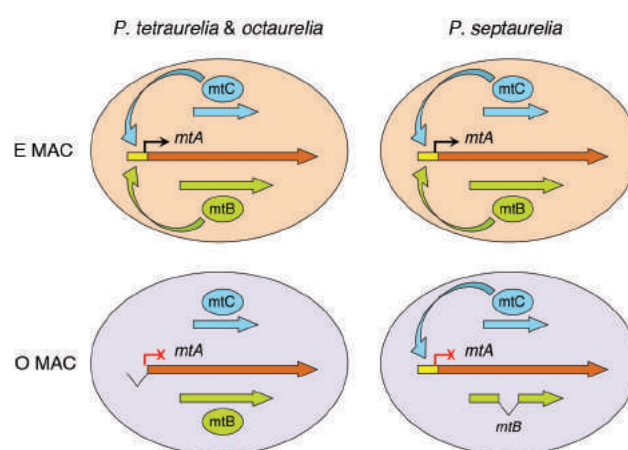


Figure 6



ANNEXE 2

RNA-guided DNA rearrangements in ciliates: Is the best genome defence a good offence?

Robert S. Coyne^{*1}, Maoussi Lhuillier-Akakpo^{†1} and Sandra Duhaucourt^{†2}

^{*}J. Craig Venter Institute, Rockville, MD 20850, U.S.A., and [†]Institut Jacques Monod, CNRS, UMR 7592, Université Paris Diderot, Sorbonne Paris Cité, F-75205 Paris, France

Genomes, like crazy patchwork quilts, are stitched together over evolutionary time from diverse elements, including some unwelcome invaders. To deal with parasitic mobile elements, most eukaryotes employ a genome self-defensive manoeuvre to recognise and silence such elements by homology-dependent interactions with RNA–protein complexes that alter chromatin. Ciliated protozoa employ more ‘offensive’ tactics by actually unstitching and reassembling their somatic genomes at every sexual generation to eliminate transposons and their remnants, using as patterns the maternal genomes that were rearranged in the previous cycle. Genetic and genomic studies of the distant relatives *Paramecium* and *Tetrahymena* have begun to reveal how such events are carried out with remarkable precision. Whole genome, non-coding transcripts from the maternal genome are compared with transcripts from the zygotic genome that are processed through an RNA interference (RNAi)-related process. Sequences found only in the latter are targeted for elimination by the resulting short ‘scanRNAs’ in many thousand DNA splicing reactions initiated by a domesticated transposase. The involvement of widely conserved mechanisms and protein factors clearly shows the relatedness of these phenomena to RNAi-mediated heterochromatic gene silencing. Such malleability of the genome on a generational time scale also has profound evolutionary implications, possibly including the epigenetic inheritance of acquired adaptive traits.

Introduction

RNA is a remarkably versatile molecule that may well have served as the original genetic template as well as replicator in a primordial ‘RNA World’ (Cech, 1993). As early life forms grew in complexity, RNA undoubtedly took on additional roles in cellular regulation and defence. Recent studies have begun to reveal the widespread and diverse functions of small RNAs (sRNAs) in modern organisms, particularly in controlling gene expression and providing defence against parasites, including genomic para-

sites (Girard and Hannon, 2008; Li and Liu, 2011; Siomi et al., 2011). For example, in bacteria and archaea, crRNAs produced by Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) loci guide destruction of invading genetic material such as phage DNA (Karginov and Hannon, 2010). In eukaryotes, mobilisation of transposable genetic elements is held in check by a variety of mechanisms involving sRNAs, including both transcriptional and post-transcriptional gene silencing (TGS and PTGS) (Rigal and Mathieu, 2011; Volpe and Martienssen, 2011). TGS of repetitive genomic regions involves their homology-dependent recognition, mediated by sRNAs, and the deposition of specific chromatin modifications, in particular methylation of both DNA and histones. This results in the formation of heterochromatin, which not only silences mobile DNA but also modulates cellular gene expression and functions in chromosome segregation (Creamer and Partridge, 2011; Siomi and Siomi, 2008). Thus, mechanisms for genome defence have

¹Robert S. Coyne and Maoussi Lhuillier-Akakpo contributed equally to this work.

²To whom correspondence should be addressed (email duhaucourt@ijm.univ-paris-diderot.fr).

Key words: Ciliated protozoa, Domesticated transposase, Epigenetic inheritance, Heterochromatin, Programmed genome rearrangement, Small RNAs.

Abbreviations used: Cbs, chromosome breakage sequence; GFP, green fluorescent protein; IES, internal eliminated sequence; NHEJ, non-homologous end joining; PHD, plant homeodomain; PTGS, post-transcriptional gene silencing; RNAi, RNA interference; RT-PCR, reverse transcription-PCR; scnRNA, scanRNA; sRNA, small RNA; TGS, transcriptional gene silencing.

become intertwined with gene regulation, heredity and ultimately the evolution of genome structure, as well exemplified by the case of the ciliated protozoa.

Taking genome defensive adaptations a step further than simply silencing invasive DNA, the ciliated protozoa use a class of sRNAs to guide genome-wide programmed DNA rearrangements that physically eliminate mobile elements from their somatic nuclei—in a sense, taking a more ‘offensive’ approach to genome defence (Chalker and Yao, 2011; Duharcourt et al., 2009; Schoeberl and Mochizuki 2011). Detailed studies of this process have shown how, mechanistically, it shares many conserved features with TGS and, likewise, with more limited forms of programmed DNA rearrangements, such as vertebrate immunoglobulin gene V(D)J recombination. Over the course of evolutionary history, fundamental aspects of the nature and extent of genome rearrangement have undergone considerable diversification amongst the ciliates, in some cases leading to bizarre genome structures [as for example amongst the spirotrichous ciliates, which carry essentially every gene on a separate somatic chromosome (Jonsson et al., 2009)]. Experimental studies have so far been limited to a few ciliate species. Here, we will focus primarily on the two most well studied, *Paramecium tetraurelia* and *Tetrahymena thermophila*. Although these species appear superficially similar to the eye, their lineages diverged about the same time as those of humans and insects.

One cell, two divergent nuclei

One of the defining characteristics of all ciliates is nuclear dimorphism. Although unicellular, they carry two structurally and functionally distinct types of nuclei, different species having one or more of each. Diploid micronuclei act as the lineage’s germline, engaging in meiosis to produce haploid gametic nuclei that are exchanged during the sexual process of conjugation (or fuse in the self-fertilisation process of autogamy common in *Paramecium*, but not *Tetrahymena*). For reasons that will shortly be clear, gene expression is limited to the polyploid macronuclei, which are destroyed during conjugation or autogamy, thus fulfilling the role played by somatic nuclei in metazoans. Following the sexual process, a new macronucleus and a new micronucleus (or more) differentiate from mitotic copies of the new zygotic nucleus.

Despite originating from identical daughter nuclei, macronuclei and micronuclei differ dramatically in their genome content and organisation due to endoreplication and the genome-wide programmed DNA rearrangement events referred to above, which occur during macronuclear differentiation. In both *Paramecium* and *Tetrahymena*, the principal rearrangement events are chromosome fragmentation and internal DNA elimination, but the nature of these events differs substantially between the two species (see Figure 1).

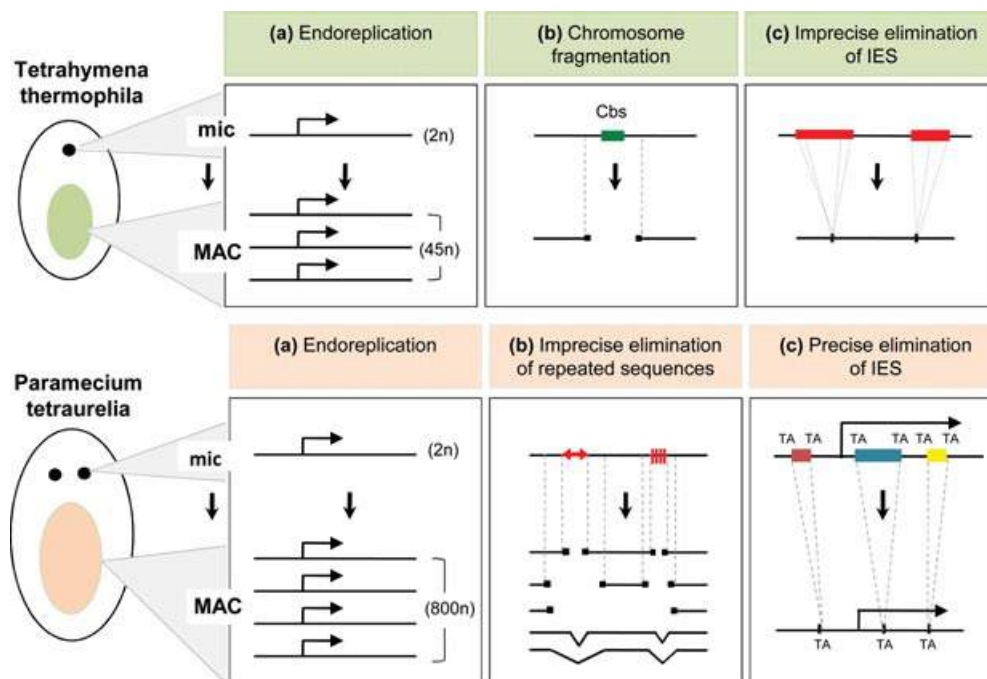
Cutting chromosomes down to size

The most straightforward event to describe is *Tetrahymena* chromosome fragmentation, a highly reproducible process by which the five micronuclear chromosomes are broken into 181 macronuclear chromosomes, ranging from 20 kb to over 3 Mb in size. The process is guided by the necessary and sufficient 15 bp chromosome breakage sequence (Cbs) (Yao et al., 1990), implying the existence of a sequence-specific endonuclease, although one has yet to be identified (see further discussion below). Only minimal loss of genetic material surrounding Cbs occurs, and the free ends are rapidly capped by telomerase (Fan and Yao, 1996). In contrast, *Paramecium* chromosome fragmentation is a more variable affair, involving the imprecise excision of micronuclear-limited repetitive sequences, such as mini-satellites and transposons, followed by the *de novo* addition of telomeres to the flanking sequences at variable (by up to 20 kb) endpoints. In some cases, however, the breakpoints are not capped by telomeres but rather joined together (Le Mouel et al., 2003). Sequencing of such junctions suggests a mechanistic similarity to the precise excision of shorter internal elements described below. Thus, *Paramecium* may lack a separate pathway for chromosome fragmentation similar to *Tetrahymena*’s Cbs-mediated process.

The precise amount of DNA eliminated from regions of *Paramecium* chromosome breakage sites is unknown, but the end result is that the macronuclear genome is almost entirely free of repetitive sequences. Because polyploid ciliate macronuclei carry thousands of chromosomes, their orderly segregation at each vegetative cell division would seem nearly impossible and, in fact, macronuclear chromosomes appear to lack active centromeres (Cervantes et al.,

Figure 1 | Nuclear dimorphism and DNA rearrangements in the ciliates *Paramecium tetraurelia* and *Tetrahymena thermophila*

Like all ciliates, these species contain two distinct types of nuclei that develop from the same zygotic nucleus at each sexual cycle: one (in *Tetrahymena*) or two (in *Paramecium*) diploid germinal micronuclei and one polyploid somatic macronucleus. Three major types of genomic events occur during macronuclear development. In *Tetrahymena*: (a) the macronuclear genome is endoreplicated to contain around 45 copies of each chromosome; (b) the 15-bp chromosome breakage sequence (Cbs) promotes chromosome fragmentation, which is followed by *de novo* telomere addition; and (c) about 6,000 moderately repetitive internal eliminated sequences (IES) present in intergenic and intronic regions are somewhat imprecisely eliminated. In *Paramecium*, analogous events occur, but the details differ in several regards: (a) many cycles of endoreplication generate a highly polyploid macronuclear genome containing 800 copies of each chromosome; (b) imprecise elimination of repeated sequences (transposon and minisatellite) leads either to chromosome fragmentation and *de novo* telomere addition or to internal deletions if the flanking regions are religated; and (c) unique TA-limited IES are precisely excised from genic and intergenic regions, restoring open reading frames and allowing macronuclear gene expression.



2006; Cui and Gorovsky, 2006). Therefore, instead of an orderly segregation, the macronucleus simply splits in half by an amitotic process. Despite this imprecise division, it appears that both species exert some form of copy number control to maintain roughly equal chromosome representation (apart from the highly amplified rRNA-encoding gene, which occupies its own chromosome) (Berger and Schmidt, 1978; Duret et al., 2008; Eisen et al., 2006; Larson et al., 1991). Nevertheless, at least in *Tetrahymena*, the inequivalent partitioning of variant alleles leads stochastically to the vegetative descendants of a single sexual progeny having highly variable geno-

types (a process known as phenotypic assortment) (Doerder et al., 1992; Orias and Flacks, 1975). This may dramatically influence the ability of some ciliates to rapidly adapt to environmental variability.

Splicing out unwanted DNA

Compared with chromosome fragmentation, greater experimental attention has been devoted to understanding the elimination of DNA sequences referred to in both *Paramecium* and *Tetrahymena* as internal eliminated sequences (IESs), although the characteristics of these elements differ greatly between the

two species. *Paramecium* IESs are unique sequence elements between 26 and 882 bp in length and about 45,000 in number that are precisely excised from the macronuclear genome, of necessity because many are found within gene-coding regions (Gratias and Betermier, 2001) (O. Arnaiz, personal communication). In contrast, the approximately 6,000 *Tetrahymena* IESs contain repetitive sequence elements, measured from less than 0.5 to over 20 kb in length and are somewhat imprecisely excised (Chalker and Yao, 2011). Nearly all are found in intergenic regions or introns (Fass et al., 2011). Given that the common ancestor of these two species almost certainly had IESs of some kind, it is curious how they have diverged so greatly, a question for which we need to consider in greater detail IES structures and their mechanisms of excision.

All *Paramecium* IESs are bound by the dinucleotide sequence TA, one copy of which is retained in the macronucleus. This dinucleotide is often part of a short inverted repeat extending into the IES and into the macronuclear destined DNA (Betermier, 2004; Klobutcher and Herrick, 1995). Single base substitutions within the TA dinucleotide have been shown to render the IES excision defective (Mayer and Forney, 1999; Ruiz et al., 2000). Furthermore, a loosely conserved consensus sequence can be defined, beginning with the TA and extending into the IES, as 5'-TAYAGYNR-3'. A related consensus can be found at the termini of IESs from the distantly related ciliate *Euplotes crassus*. Remarkably, these sequences resemble the terminal consensus for transposons of the Tc1/Mariner superfamily (Klobutcher and Herrick, 1995). This correspondence led to the proposal that the IESs of both ciliate species are evolutionary remnants of mobile elements that have degenerated beyond sequence recognition, apart from their termini (Klobutcher and Herrick, 1997). In this scenario, the 'excisase' would presumably be related to a transposase either retained in one or more intact transposons or 'domesticated' by the organism to serve its genome scrubbing purposes. This general hypothesis has received considerable support from investigations into the excision mechanism, but with a surprising twist. Although the *Paramecium* micronuclear genome has been found to contain at least fragments of complete Tc1/Mariner transposons, the role of mediating their excision, as well as that of their presumed remnants (IESs) has been as-

sumed by the transposase of a different class of mobile element.

A domesticated piggy

Tc1/Mariner transposases mobilise DNA by a cut-and-paste mechanism initiated by staggered cleavage flanking a TA dinucleotide, leaving 2-base, 3' overhangs (Plasterk et al., 1999). Unexpectedly, the free ends generated during *Paramecium* IES excision have 4-base, 5' overhangs centred on the TA, implying cleavage by a different endonuclease family (Gratias and Betermier, 2003). A candidate was identified in the genome sequence by similarity to the *piggyBac* transposase family and subsequently named PiggyMac, for its role in macronuclear differentiation (Baudry et al., 2009). *In vitro*, the *piggyBac* transposase catalyses double-strand cleavage at *piggyBac* transposon ends, and the resulting double-strand breaks exhibit 4-base, 5' overhangs (Mitra et al., 2008). The predicted PiggyMac protein contains the conserved DDD (Asp residues) catalytic triad necessary for endonuclease function, as well as a plant homeodomain (PHD) finger that might help mediate IES recognition (see below). Several lines of evidence support the conclusion that PiggyMac is the enzyme responsible for initiating IES excision. The gene is single copy and no longer contained on a transposon; thus, it has been thoroughly domesticated. Apparently, this event occurred prior to the split between *Paramecium* and *Tetrahymena*, which also depends on a *piggyBac* relative for internal DNA elimination (see below). Independent domestication of *piggyBac*-derived sequences has been reported in several other eukaryotic genomes (Sarkar et al., 2003).

As noted, many *Paramecium* IESs are removed from gene-coding regions, so their excision must be precise to the nucleotide. This precision appears to be achieved by a concerted 'cut-and-close' mechanism. Cleavage events at the two ends of an IES are tightly coupled and the free ends held together, perhaps in part by base pairing of the central palindromic TA dinucleotides on each strand, until resection (if necessary) of unpaired single overhanging nucleotides, repair synthesis and ligation are completed (Gratias and Betermier, 2003; Gratias et al., 2008). These events are dependent on components of the non-homologous end joining (NHEJ) pathway, including homologues of Lig4p and XRCC4 (Kapusta et al., 2011).

In many respects, IES excision resembles V(D)J joining during vertebrate immunoglobulin gene rearrangement. This process is also initiated by an endonuclease derived from a domesticated transposase (Kapitonov and Jurka, 2005) and dependent on NHEJ for healing of the broken ends (Lieber et al., 2004). In contrast to *Paramecium* IES excision, the microheterogeneity introduced during this healing process (which contributes to immunoglobulin diversity) can be explained by differences in the free ends to be joined, which are blunt, and in their processing.

***Tetrahymena* splices differently**

In *Tetrahymena*, the full nature of micronuclear-limited DNA sequences is unclear. Early experiments using DNA re-association kinetics indicated that most of these sequences are moderately repetitive (100 copies per genome or fewer) and that they constitute about 10–20% of the micronuclear genome (Yao and Gorovsky, 1974). By comparing random micronuclear genomic regions to their macronuclear counterparts, it was estimated that about 6,000 IESs exist and that their typical size is in the 1–2 kb range (Yao et al., 1984). This approximate number is supported by a recent whole micronuclear genome sequencing survey (Fass et al., 2011), but preliminary whole genome assembly indicates the micronuclear genome to be 50% greater in size than that of the macronucleus (<http://www.broadinstitute.org/annotation/genome/Tetrahymena/GenomeStats.html>). In addition, analysis of several repetitive elements has shown them to be excised as parts of larger blocks of micronuclear-limited DNA, at least up to 22 kb in size. Some of these elements are clearly related to retrotransposons (Wuitschick et al., 2002). Representatives of several other transposon families, including Tc1/Mariner, have also been detected (Eisen et al., 2006). Because of the difficulties inherent in assembling such repetitive genomes, a full picture of the *Tetrahymena* germline may take some time to emerge.

Most studies of *Tetrahymena* IES excision have focused on a limited number of 0.6–3 kb elements, typified by the so-called M and R elements. The ends of such IESs are characterised by short (2–8 bp) direct repeats, but no consensus sequence can be defined, and indeed, excision is often somewhat imprecise in its

junctions (Chalker and Yao, 2011). As with *Paramecium*, the initiating step is a staggered cut, leaving 4 base, 5' overhangs, but breakage at the two ends does not appear to be coupled; instead, it has been proposed that the free macronuclear-destined end initiates a transposition-like transesterification reaction to join itself to the macronuclear-destined sequences at the other end of the IES (Saveliev and Cox, 1996). As with *Paramecium*, there is strong evidence that the initiating cut is carried out by a domesticated piggyBac homologue (Cheng et al., 2010). The gene product, Tpb2p, has been shown *in vitro* to exhibit a double-strand endonuclease activity consistent with this role. Tpb2p shows a sequence preference in its cutting site similar to PiggyBac from other organisms and also cleaves a synthetic *Tetrahymena* IES junction sequence at the same site previously found to initiate excision *in vivo*. Silencing of the TPB2 gene severely reduces the efficiency of both IES excision and chromosome fragmentation (residual activity may be due to incomplete silencing). Thus, Tpb2p could potentially be the nuclease responsible for chromosome breakage, if coupled with a sequence-specific factor that recognises Cbs. Alternatively, chromosome breakage may depend for its completion on chromatin or other events that are disrupted in the TPB2-silenced cells; this latter interpretation is favoured by knock-out or knock-down studies of several other DNA elimination factors that also indicate effects on chromosome breakage (Malone et al., 2005; Matsuda et al. 2010; Mochizuki et al. 2002; Motl and Chalker, 2011; Nikiforov et al., 1999).

By detailed deletion analyses, *cis*-acting sequences were identified in the macronuclear-destined regions flanking M, R and other elements that define excision endpoints, but they are not conserved between elements and can often be only roughly mapped (Chalker and Yao, 2011). Intriguingly, regions within the IESs themselves are also required for efficient excision, but again, these are not widely conserved or easily mapped (Chalker and Yao, 2011). Likewise, the short consensus at the ends of *Paramecium* IESs is not up to accounting for the specificity of 45,000 precise excision events (O. Arnaiz, personal communication); in fact, this consensus is not sufficient to define excision sites, although it may help determine excision boundaries and efficiency (Duharcourt et al., 2009). So, how is this massive genome rearrangement carried out with such astonishing efficiency and precision?

Ultimately, this puzzle gave way to an explanation based on homology-dependent epigenetic communication between the maternal macronucleus and the newly developing macronucleus.

Epigenetic regulation of genome rearrangements: Beyond Mendel

A long history of studies on ciliates has shown that the maternal somatic genome can determine inheritance of traits by the succeeding sexual generation, in a manner independent of the germline genotype. Over time, it was recognised that this epigenetic influence is mediated by the regulation of genome rearrangement events. The first such case to be understood at a molecular level was the d48 epimutant of *Paramecium* (Epstein and Forney, 1984). This cell line fails to express the A surface antigen gene because of a terminal chromosomal deletion in the macronucleus that is inherited faithfully from one generation to the next, despite the fact that the germline copy of the gene is entirely wild type (Forney and Blackburn, 1988). Thus, the trait follows a cytoplasmic rather than a Mendelian pattern of heredity.

Not only terminal deletions, but also IES excision patterns have been shown to be maternally controlled and can indeed be experimentally manipulated by genetic transformation of the maternal macronucleus. In both *Paramecium* and *Tetrahymena*, introduction of a particular IES sequence into the maternal macronucleus leads to the retention of the homologous germline sequence in the new macronucleus of sexual progeny, whereas the excision of all other IESs occurs normally (Chalker and Yao, 1996; Duhaucourt et al., 1995). This sequence-specific maternal inhibition of IES excision was observed for a third of the IESs tested in *Paramecium*, which have therefore been called maternally controlled IESs (mcIESs), although the effect varies quantitatively amongst them (Duhaucourt et al., 1998). In some instances, this inhibition results in variant cell lines in which an IES is stably maintained in the macronuclear genome from one sexual generation to the next (IES + cell line), even though the germline genome is entirely wild type, analogous to the d48 epimutant.

Alternative excision profiles are thus epigenetically inherited and appear to depend on the content of the maternal macronucleus. This suggests that what defines an eliminated element is not primarily its own sequence or its flanking sequences but rather

its absence from the maternal somatic genome. In support of this idea, a foreign sequence (neomycin resistance gene) introduced only into the germline genome of *Tetrahymena* is frequently eliminated from the newly developed macronucleus (Yao et al., 2003). Interestingly, the frequency of elimination is dosage dependent, increasing with the number of germline copies. These results imply the production of a *trans*-acting substance from the invasive gene sequences (Liu et al., 2005).

Altogether, these observations led to the proposal that the recognition of eliminated elements is based on a *trans*-nuclear comparison between the germline genome to be rearranged with the previously rearranged genome contained in the maternal macronucleus, which is still present in the cytoplasm during the development of the new one (see Figure 2). Essentially, any sequence not found in the maternal macronucleus is targeted for elimination in the developing macronucleus. To account for the high sequence specificity of this genome comparison, it was hypothesized to be mediated by pairing interactions between homologous nucleic acids.

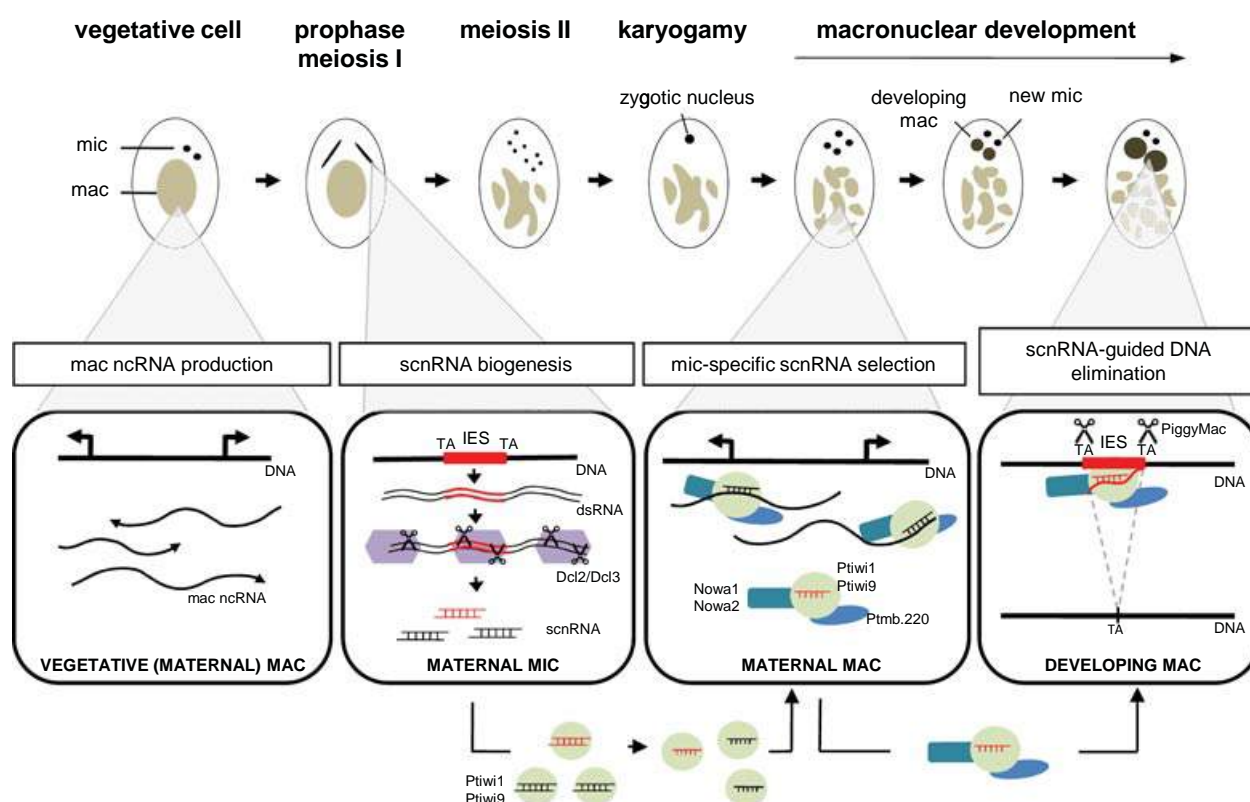
RNA-guided programmed genome rearrangements

Maternal germline scnRNAs

A major breakthrough in the field came from the discovery of a developmentally specific RNA interference (RNAi)-related pathway required for programmed genome rearrangements. In *Tetrahymena* and *Paramecium*, scanRNAs (scnRNAs), approximately 28 and 25 nt in length, respectively, are exclusively accumulated during sexual events (Lepere et al., 2009; Mochizuki et al., 2002;). These molecules are produced from the germline micronuclear genome during meiosis by an RNAi-related pathway. The micronuclear genome is transcriptionally silent for most of the life cycle, but, at least in *Tetrahymena*, it is active during prophase of meiosis I (Martindale et al. 1985; Sugai and Hiwatashi, 1974), the only stage during which components of the transcriptional machinery localise to the micronucleus (Mochizuki and Gorovsky, 2004b; Stargell and Gorovsky, 1994). At this time, bidirectional transcripts are synthesized (Chalker and Yao, 2001), likely forming double-stranded RNA molecules that are processed into scnRNA duplexes by the Dicer-like

Figure 2 | The “genome-scanning” model, as envisioned in *Paramecium* (similar events and factors are involved in *Tetrahymena*)

The progression of nuclear developmental events is represented above, and below, the corresponding steps of the scanning model and key factors involved. (i) Macronuclear non-coding transcription. Non-coding transcripts are constitutively produced from both strands of the whole macronuclear genome. (ii) Biogenesis of scnRNAs. Early in meiosis, transcription of the whole micronuclear genome also occurs. 25-nt scnRNA duplexes are produced by the ribonuclease III activity of the Dicer-like proteins Dcl2 and Dcl3. These duplexes are then transported to the cytoplasm where they are loaded onto Ptiwi01–Ptiwi09 complexes, leading to removal of the “passenger” strand. (iii) Micronuclear-specific scnRNA selection. Ptiwi–scnRNA complexes are transported to the maternal macronucleus. Homologous pairing between scnRNAs and macronuclear non-coding transcripts, enhanced by the putative helicase Ptmb.220 and assisted by the putative RNA binding proteins Nowa1–Nowa2, leads to selection of micronuclear-specific scnRNAs that are then transported to the developing macronucleus. (iv) DNA elimination. Recognition of scnRNA-homologous sequences in the developing macronuclear genome targets recruitment of the endonuclease PiggyMac, allowing elimination of micronuclear-specific sequences.



protein Dcl1p (Malone et al., 2005; Mochizuki and Gorovsky, 2005). In *Paramecium*, production of scnRNAs requires two Dicer-like proteins, Dcl2 and Dcl3 (Lepere et al., 2009). Sequencing of *Paramecium* scnRNAs has revealed a prominent 5'-UNG signature and a minor fraction showing the complementary signature at positions 21–23, indicating that scnRNAs are cleaved from double-stranded precursors as duplexes with 2-nt 3' overhangs at both ends, followed by preferential stabilisation of the 5'-UNG

(Lepere et al., 2009). *Tetrahymena* scnRNAs also display a strong 5' U bias (Lee and Collins, 2006).

The scnRNA population is highly complex and does not exclusively correspond to micronuclear-limited sequences. Indeed, scnRNAs are produced from a large fraction of the germline genome including both strands of exons, introns, intergenic regions, transposons and IESs (Lee and Collins, 2006; Lepere et al., 2009). The data obtained so far are consistent with the idea that the entire micronuclear genome is

transcribed and produces scnRNAs. In *Tetrahymena*, loss of Dcl1p results in the accumulation of a heterogeneous population of non-coding RNAs, likely the scnRNA precursors. Knock-out of *Tetrahymena* Dcl1p or knock-down of both *Paramecium* Dcl2 and Dcl3 results in the disappearance of scnRNAs, DNA elimination defects and eventually death in the sexual progeny (Malone et al., 2005; Mochizuki and Gorovsky, 2005; Lepere et al., 2009). Thus, Dicer-dependent scnRNAs are required for the elimination of germline-limited sequences both in *Paramecium* and *Tetrahymena*.

These scnRNAs are associated with members of the Argonaute family, the Piwi proteins. The scnRNA-associated Piwi protein Twilp in *Tetrahymena* is required for scnRNA accumulation (Mochizuki et al., 2002). In *Paramecium*, two Piwi proteins, Ptiwi01 and Ptiwi09, were shown to be the functional counterparts of Twilp (Bouhouche et al., 2011). *Tetrahymena* scnRNAs are 2'-O-methylated at their 3' termini by the RNA methyltransferase Hen1p, which is required for scnRNA stability and efficient DNA elimination (Kurth and Mochizuki, 2009). All the genes involved in the metabolism of scnRNAs (production, accumulation and stability) are necessary for the elimination of germline-limited sequences in the developing macronucleus, demonstrating the crucial role that scnRNAs play in this process. The most direct evidence that scnRNAs can target the elimination of an IES in the developing macronucleus was obtained by microinjection experiments with 25-nt synthetic RNA duplexes that mimic the structure of endogenous scnRNAs (Lepere et al., 2008).

There are strong similarities between scnRNAs and metazoan piRNAs. Structurally, both are longer than siRNAs, have a strong 5' U bias and are methylated by the Hen1 protein. Functionally, both are required for sexual reproduction and for the control of transposable elements (or their presumed remnants, IESs)—piRNAs through epigenetic silencing and scnRNAs through epigenetic elimination. Yet, scnRNAs also show some differences with piRNAs. Unlike the latter, scnRNAs act during the formation of the somatic genome post-zygotically and are not known to contribute to silencing of the germline genome. In contrast to scnRNAs, Dicer-like proteins do not seem to be involved in the biogenesis of piRNAs (Li and Liu, 2011). Conversely, scnRNAs show no evidence for a 'ping-pong' mechanism of amplification involving

the endoribonucleolytic ('slicer') activity of the Piwi proteins (Brennecke et al., 2007; Gunawardane et al., 2007).

Maternal somatic non-coding transcripts

So, if scnRNAs are produced from the entire micronuclear genome, how do they target for elimination only those sequences absent from the maternal macronucleus? In contrast to scnRNAs, which appear to promote DNA elimination of homologous sequences, another class of non-coding transcripts, produced by the maternal somatic genome, appears to protect homologous sequences against elimination. Examples of non-coding transcription for both coding and intergenic regions have been documented in *Paramecium* and *Tetrahymena* and suggest that the whole macronuclear genome is transcribed (Aronica et al., 2008; Chalker and Yao, 2001; Chalker et al., 2005; Lee and Collins, 2006; Lepere et al., 2008). In *Paramecium*, reverse transcription-PCR experiments have revealed that constitutive non-coding transcripts are produced by the macronucleus at low levels on both strands. Targeted degradation of these transcripts by RNAi triggers deletion of their homologous DNA sequences in the developing macronucleus (Lepere et al., 2008). Similarly, RNAi-mediated silencing of a non-essential gene during vegetative growth (through an independent PTGS mechanism) was shown to trigger, after sexual reproduction, the deletion of the gene in the developing macronucleus (Garnier et al., 2004). Targeting a small region of the gene is sufficient to delete the entire gene, suggesting that RNAi-mediated silencing leads to the degradation of the transcripts covering the entire gene, although these are not necessarily messenger RNAs. Thus, in order to avoid elimination, homologous sequences not only need to be present in the maternal macronucleus but also to generate transcripts that counteract scnRNAs. This is analogous to the, at first glance, paradoxical situation observed in fission yeast, plants and animals in which active transcription of a certain genomic region is required for its subsequent heterochromatic silencing (Buhler and Moazed, 2007).

The 'genome scanning' model

To explain how these two classes of maternal non-coding RNAs with antagonistic functions act as epigenetic regulators of DNA elimination, a 'genome

Table 1 Factors involved in 'genome-scanning' in *Tetrahymena thermophila* and *Paramecium tetraurelia*

	<i>Tetrahymena</i>	<i>Paramecium</i>	Protein motif and/or putative function
scnRNA biogenesis (micronuclei)	Dcl1p (Malone et al., 2005; Mochizuki and Gorovsky, 2005)	Dcl2–Dcl3 (Lepere et al., 2009)	Dicer-like ribonuclease III
Genome scanning (maternal macronucleus)	Twilp (Mochizuki et al., 2002)	Ptiwi01–Ptiwi09 (Bouhouche et al., 2011)	Piwi
	Giw1p (Noto et al., 2010)	ND	Piwi binding
	Hen1p (Kurth and Mochizuki, 2009)	ND	2'-O-RNA methyltransferase
	Wag1–ConjB (Bednenko et al., 2009)	Nowa1–Nowa2 (Nowacki et al., 2005)	GW repeats
	Ema1p (Aronica et al., 2008)	Ptmb.220 (Nowak et al., 2011)	RNA helicase

ND: not determined.

scanning' model was proposed (see Figure 2, Table 1). According to this model, scnRNAs would scan the maternal somatic genome for homologous sequences through base pairing interactions with nascent non-coding transcripts. This scanning procedure would allow the selection of micronuclear-specific scnRNAs (those that fail to find homologous partners in the maternal genome) that would specifically guide the deletion of germline-limited sequences in the developing macronucleus. Pairing interactions between deletion-promoting germline scnRNAs and protective somatic non-coding transcripts would thus mediate a natural genomic subtraction to allow elimination of all sequences absent from the maternal somatic macronucleus.

The scnRNA-associated Piwi proteins, Twilp and Ptiwi09, are found in the maternal macronucleus before the new macronucleus has developed (Bouhouche et al., 2011; Mochizuki et al., 2002), supporting the idea that the scanning process takes place within the maternal macronucleus. After production in the micronuclei at meiosis, scnRNA duplexes appear to be exported to the cytoplasm and loaded onto scnRNA-associated Piwi proteins (Bouhouche et al., 2011; Mochizuki et al., 2002; Noto et al., 2010). The *Tetrahymena* Piwi protein Twilp, like most Argonaute proteins, has a slicer activity that cleaves one strand of the scnRNA duplex, leading to removal of the 'passenger' strand. This slicer activity has been shown to be required for the import of Twilp into the maternal macronucleus (Noto et al., 2010). Twilp complexed with single-stranded scnRNAs are likely transported to the parental macronucleus, with the help of Giw1p, a Twilp-binding protein (Noto et al., 2010). These scnRNAs are then 2'-O-methylated by

Hen1p, which is localised to the maternal macronucleus (Kurth and Mochizuki, 2009).

Pairing interactions between scnRNAs and non-coding transcripts in the maternal macronucleus are believed to allow selection of micronuclear-limited scnRNAs from the initial, highly complex population. Those scnRNAs that correspond to macronuclear sequences would hybridise with complementary macronuclear non-coding transcripts, and this pairing interaction would somehow result in their inactivation. In contrast, those scnRNAs that correspond to germline-limited sequences would by definition not find any complementary sequences in the maternal macronucleus and therefore be free to be exported to the zygotic macronucleus, once it develops, where they would trigger the elimination of homologous sequences.

The mechanisms by which macronuclear scnRNAs are inactivated have not yet been elucidated. One possibility would be that scnRNAs are sequestered by the non-coding macronuclear transcripts to which they pair, which would act as molecular sponges. It could also be that scnRNA pairing interactions induce their degradation. In favour of this latter hypothesis, the scnRNA population appears to be enriched for micronuclear-specific sequences as the sexual process proceeds (Mochizuki and Gorovsky, 2004a; Lee and Collins, 2006; Lepere et al., 2009).

A number of essential protein factors in this process have been identified. In *Tetrahymena*, the Twilp-binding, putative RNA helicase Ema1p is essential for the interaction between scnRNAs, Twilp complexes and nascent maternal macronuclear transcripts and appears to function in the selective degradation of at least some macronuclear scnRNAs

(Aronica et al., 2008). Dissociation of Twi1p from scnRNAs paired with maternal ncRNAs might result in scnRNA degradation and recycling of Twi1p (Aronica et al., 2008). In *Paramecium*, the putative RNA helicase Ptmb.220p is required for DNA rearrangements and might play a role similar to its *Tetrahymena* counterpart Ema1p (Nowak et al., 2011). The RNA-binding, GW repeat-containing Nowa1 and 2 proteins in *Paramecium*, which are required for the excision of mcIESs and transposons, localise to the parental macronucleus and have been suggested to participate in the scanning process (Nowacki et al., 2005), but further work is needed to precisely determine their role. In *Tetrahymena*, two related GW (Gly-Trp residues) repeat-containing proteins, ConjB and Wag1, have also been shown to localise to the maternal macronucleus and to be essential for DNA elimination. Interestingly, they interact with Twi1p and appear to function in scnRNA selection (Bednenko et al., 2009). This suggests that, as in other systems, the reiterated GW motifs form an 'Ago-hook' that binds Argonaute proteins (El-Shami et al., 2007; Till et al., 2007). Green fluorescent protein (GFP) fusions have shown that a number of these protein factors, including Nowa1p, Twi1p, Ptiwi09p and Ema1p, accumulate in the maternal macronucleus before completion of meiosis of the micronucleus and are later transported to the developing macronuclei, suggesting that they are delivering scnRNAs to the developing macronucleus, where they direct DNA elimination. Next, we consider by what mechanism.

Cut here: scnRNA-mediated targeting of the endonuclease

DNA elimination in the developing macronucleus relies on the accurate recruitment of piggyBac-related endonucleases to micronuclear-specific sequences. According to the scanning model, micronuclear-specific scnRNAs that have been transported into the developing macronucleus are used to target their homologous genomic sequences and define them as sequences to be eliminated. There are several clues as to how this is achieved, but the full mechanism has not yet been determined. The evidence also suggests significant differences between *Tetrahymena* and *Paramecium*.

In *Tetrahymena*, two heterochromatin-specific marks, trimethylation of histone H3 on lysine 9 (H3K9me3) and lysine 27 (H3K27me3), are asso-

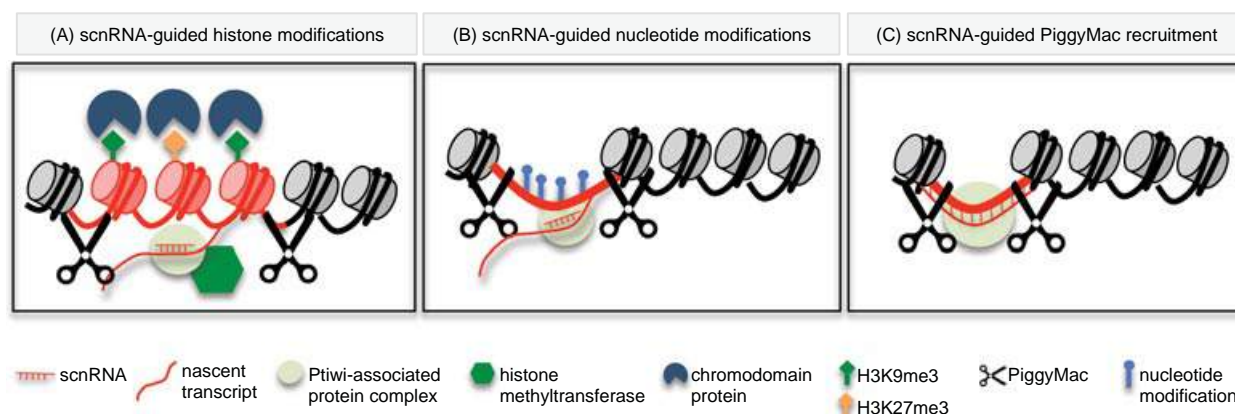
ciated with IES sequences around the time of their excision (Taverna et al., 2002; Liu et al., 2007). These modifications are directly required for DNA elimination, as mutations in the H3 amino acid sequence preventing K9 or K27 methylation lead to IES retention (Liu et al., 2004, 2007). The SET domain histone methyltransferase Ezl1p, a homologue to *Drosophila* Enhancer of zeste, regulates the deposition of these marks during development and is also required for IES elimination (Liu et al., 2007). Furthermore, Twi1p silencing abolishes H3K27 and H3K9 methylation in the developing macronucleus, providing evidence that developmental histone methylation by Ezl1p is scnRNA-dependent (Liu et al., 2004, 2007; Malone et al., 2005). IES-associated H3K9me3 and H3K27me3 are then bound by the conjugation-specific chromodomain proteins Pdd1p and Pdd3p (Pdd1p can bind both marks, whereas Pdd3p recognises only H3K9me3), triggering the formation of higher-order heterochromatic structures at the nuclear periphery (Smothers et al., 1997; Taverna et al., 2002). These structures are essential for DNA elimination and might recruit the endonuclease Tpb2p for IES elimination (Cheng et al., 2010). It is striking that, despite concerted efforts, IES excision activity has yet to be demonstrated in an *in vitro* system, perhaps indicative of precise chromatin architecture requirements for this reaction.

How scnRNAs guide DNA elimination in *Paramecium* remains to be discovered. In parallel with *Tetrahymena*, the targeting of eliminated sequences could occur via scnRNA-induced deposition of as yet uncharacterised heterochromatin modifications on homologous sequences. These marks would then be bound by specific chromodomain proteins allowing recruitment of PiggyMac at each boundary of the targeted sequence (Figure 3A). Several putative SET domain histone methyltransferases and chromodomain proteins exist in the *Paramecium* genome and show developmentally up-regulated expression during sexual events (Arnaiz et al., 2010; Aury et al., 2006). Further work is needed to evaluate the role of histone methylation in this species.

There are striking similarities between scnRNA-directed DNA elimination and RNAi-dependent heterochromatin formation in other eukaryotes. In *Schizosaccharomyces pombe* for instance, pericentromeric repeats are transcribed from both DNA strands, generating double-stranded RNAs that are cleaved into

Figure 3 | Three alternative hypothetical pathways for scnRNA-mediated recruitment of the endonuclease in *Paramecium*

(A) scnRNA-guided histone modifications. Ptiwi-associated scnRNAs hybridise with homologous nascent zygotic transcripts and target deposition of heterochromatic histone methylation marks (H3K9me3 and H3K27me3) by recruitment of histone methyltransferases. These marks are recognised by chromodomain proteins, and heterochromatinisation of the targeted sequence ensues. PiggyMac, perhaps recruited by chromodomain proteins, induces double-strand breaks at the borders of the targeted sequence. (B) scnRNA-guided nucleotide modifications. Ptiwi-associated scnRNAs target deposition of nucleotide modifications that allow PiggyMac recruitment and elimination of the targeted sequence. (C) scnRNA-guided PiggyMac recruitment. For the shortest IESs, pairing of scnRNAs ensures the precise guiding of DNA cleavage next to the TA by direct recruitment of the PiggyMac endonuclease.



siRNAs by the Dicer protein Dcr1. Loaded into the Ago1 (Argonaute)-containing RNA-induced initiation of transcriptional silencing (RITS) complex, siRNAs target the dimethylation of H3K9 by the histone methyltransferase Clr4 at their homologous loci. This histone mark is recognised and bound by the chromodomain protein Swi6, which promotes heterochromatinisation of the pericentromeric regions (Lejeune et al., 2011). In this pathway, transcription of siRNA targets is required for the assembly of silent chromatin. Moreover, it has been shown that Ago1 associates with target transcripts and RNA polymerase II, suggesting that RNA–RNA interactions between siRNAs and nascent target transcripts are required to promote locus-specific chromatin modifications (Buhler et al., 2006). Similarly, *Tetrahymena* Twilp, loaded with micronuclear-specific scnRNAs, enters the developing macronucleus, where it interacts with chromatin as well as with non-coding transcripts (implying that these transcripts are nascent). These Twilp associations are dependent on the putative RNA helicase Ema1p (Aronica et al., 2008), suggesting that they are mediated by homologous base pairing between Twilp-bound sc-

nRNAs and nascent non-coding RNAs. This association could then target heterochromatic histone modifications and endonuclease recruitment to IES boundaries.

If scnRNA-directed IES targeting in ciliates uses a similar pathway as heterochromatin formation in other eukaryotes, it goes a step further with the physical elimination of targeted sequences. This might be due to the capacity of recruited endonucleases to recognise the specific chromatin structure that is produced at IESs. Indeed, PiggyMac and its *Tetrahymena* homologue Tpb2p each contain a putative PHD finger at their C-terminal ends, potentially able to bind histone modifications. Most PHD-containing proteins, including the RAG2 recombinase involved in mammalian V(D)J recombination, bind a methylated form of H3K4. This interaction is essential for RAG2 recombinase activity *in vivo* (Matthews et al., 2007). Interestingly, a few examples of PHD domains recognising H3K9 methylation have also been described. The human chromodomain protein CDH4, belonging to the nucleosome remodelling and histone deacetylase (NuRD) complex, contains two PHD fingers. Both of these domains show a

strong affinity *in vitro* for H3 peptides that are trimethylated on K9 and preferentially unmodified on K4. Binding of CDH4 on two distinct H3 tails in a single nucleosome or in adjacent nucleosomes might play an important role in chromatin remodelling dynamics (Mansfield et al. 2011; Musselman et al., 2009). Another example is the human ICBP90 factor, which requires its PHD motif to bind H3K9me3 at pericentromeric regions, regulating formation or maintenance of heterochromatin (Karagianni et al., 2008).

Although the PiggyMac PHD domain might allow it to recognise histone modifications, the hypothesis that heterochromatin formation marks *Paramecium* eliminated sequences runs up against a seeming conundrum. Genome-wide identification of PiggyMac-sensitive IESs has revealed that the vast majority are smaller than 150 bp and some are as short as 26 bp (O. Arnaiz, personal communication). How then to explain the recognition of sequences whose size is not even as long as the DNA wrapped around a single nucleosome? Access to DNA sites internal to a nucleosome might be possible through the action of ATP-dependent chromatin remodelling complexes (for example, SWI/SNF proteins) that could promote DNA 'breathing' on the nucleosome, nucleosome sliding or nucleosome removal to expose the DNA cleavage sites. However, we must consider alternative pathways, especially for recognition of short IESs.

One possibility is that *Paramecium* IESs may be defined by nucleotide modifications. Just as they might interact with histone methyltransferases, Ptiwi-scnRNA complexes could recruit DNA-modifying enzymes to mark IES DNA, allowing recognition by PiggyMac (Figure 3B). Amongst eukaryotes, DNA methylation is the most commonly found nucleotide modification. Often associated with cytosines in a CpG context, DNA methylation is involved in the silencing of many genomic regions, including repeated elements and parentally imprinted genes. In flowering plants, an RNAi-dependent DNA methylation (RdDM) pathway is responsible for transposon silencing and stable repression of endogenous genes. Transcription of RdDM-targeted loci by the plant-specific DNA-dependent RNA polymerase IV, combined with the action of an RNA-dependent RNA polymerase, generates dsRNA precursors, which are processed by the Dicer-like protein DCL3 into 24 nt siRNAs. These siRNAs are then loaded onto the Arg-

onaute protein AGO4 or AGO6 and guide methylation on homologous DNA in all cytosine sequence contexts by the *de novo* DNA methyltransferase DRM2 (Domain Rearranged Methyltransferase 2) (Zhang and Zhu, 2011). Similarly in mammals, epigenetically stable repression of transposons requires CpG methylation (Bourc'his and Bestor, 2004). In mice, *de novo* establishment of methylation of transposable elements is dependent on the piRNA pathway in the male germline (Kuramochi-Miyagawa et al., 2008). Components of the piRNA pathway are also required for *de novo* establishment of the differentially methylated region of an imprinted mouse locus (Watanabe et al., 2011).

The *Paramecium* and *Tetrahymena* genomes lack detectable cytosine methylation. However, macronuclear DNA sequences harbour N6-adenine methylation, a nucleotide modification also found in bacterial, archaeal and fungal genomes (Cummings et al., 1974; Harrison and Karrer, 1985). In *Tetrahymena*, N6-methyl-adenine modifications seem to be correlated negatively with nucleosome positioning, as the majority of these marks are found in the linker DNA (Karrer and VanNuland, 2002). Moreover, the observation of *de novo* methylation at GATC sites in the developing macronucleus at a time corresponding to micronuclear DNA elimination has led to the suggestion that N6-methyl-adenine may play a role in IES recognition (Harrison and Karrer, 1985), although there is no evidence yet supporting this hypothesis. Adenine methylation is also involved in restriction-modification systems by which it protects the bacterial genome from the action of endogenous restriction enzymes that degrade foreign, non-methylated DNA molecules (Wion and Casadesus, 2006). Interestingly, it has been suggested that adenine methylation could function in such a system in *Tetrahymena* (Blackburn et al., 1983), in this case as a protective mark against elimination. We could imagine a model, inverse to that presented in Figure 3C, in which macronuclear sequences would be defined by DNA methylation, whereas eliminated sequences would be targeted by the absence of nucleotide marks. Further investigations will determine whether adenine methylation or other forms of nucleotide modification play a role in the DNA elimination process in ciliates.

As a final alternative hypothesis to explain short IES targeting, we propose that they may be recognised without the need of any chromatin modifications.

Indeed, the PiggyMac endonuclease could be recruited to IESs by a direct interaction with scnRNA-bound Ptiwi complexes (Figure 3C). Strikingly, the size distribution of *Paramecium* IES reveals that the majority of short IES are 26 bp in length, which coincides with scnRNA length (25 nt). Furthermore, the 5'-UNG signature of the guide strand is similar to the three nucleotides at positions three to five of the degenerate consensus sequence 5'-TAYAGYNR-3' found at IES ends (Klobutcher and Herrick, 1995). This raises the possibility that pairing of scnRNAs ensures the precise guiding of DNA cleavage next to the TA by direct recruitment of the PiggyMac endonuclease (Figure 3C).

Evolutionary implications of genome rearrangement

Homology-dependent regulation of gene activity and chromatin structure by short RNAs is an ancient phenomenon (Joshua-Tor and Hannon, 2011). By coupling this mechanism to the action of a domesticated transposase, *Paramecium* and *Tetrahymena* have put their very genome structures under the sway of RNA-mediated epigenetic inheritance. This has profound evolutionary implications, ranging from environmental adaptation of small populations to fundamental alteration of both germline and somatic genomes. Our understanding of these phenomena is so far limited, most studies having been done on only a few species and in strain backgrounds of little or no genetic variability, but certain studies suggest the unusual nuclear characteristics of ciliates are linked to rapid genetic diversification (Zufall et al., 2006).

As noted above, phenotypic assortment of alleles as the result of amitotic vegetative cell division has the potential to generate wide genotypic diversity in populations descended even from single individuals. Whether this phenomenon indeed enhances the environmental adaptability of ciliates in the wild has not yet been verified but seems plausible. Genotypic variability also arises in association with IES removal and, in *Paramecium*, chromosome fragmentation. Even in matings between cells of the same inbred *Tetrahymena* strain, variable excision endpoints have been well documented (although none has been shown to affect viability). Furthermore, different inbred *Tetrahymena* strains even differ significantly in the sequence structure of individual IESs, apparently as a result

of frequent insertion/deletion events involving repetitive sequence elements (Huvos, 2004a, 2004b). In *Paramecium*, the use of a series of whole genome duplications as time markers made it possible to show that acquisition of new IESs is very likely an ongoing process (O. Arnaiz, personal communication). Such events in each species might be the result of reintegration of excised IESs at new locations. Furthermore, potentially active transposons are found in the micronuclei of each species, although their mobility may be largely suppressed by silencing, genome-wide micronuclear transcription during meiosis may afford opportunities for their proliferation.

In summary, the germline-specific regions of natural ciliate populations are likely to be highly diverse. However, as long as ciliates are able to generate a functional somatic nucleus through genome rearrangement, variability of the micronuclear genome can be tolerated. This ability depends on epigenetic recognition of 'foreign' sequences that do not belong in the new zygotic macronucleus. In *Tetrahymena*, if one mating partner has acquired a novel sequence in its micronucleus, scnRNAs homologous to this sequence will be transmitted, by cytoplasmic exchange, to the zygotic macronucleus of the other partner, allowing excision (Liu et al., 2005). *Paramecium*, however, does not engage in cytoplasmic exchange during mating, and thus novel IES acquisition could potentially result in mating incompatibility and ultimately speciation.

RNA cross-talk between nuclei or cells represents widely used mechanisms for gametic genome integrity and hybrid compatibility in plants and animals (reviewed in Bourc'his and Voinnet, 2010). In *Paramecium* and *Tetrahymena*, the precise stoichiometry between maternal RNAs produced by the germline and somatic genomes might also influence reproductive success. In case of over-production of scnRNAs or under-production of macronuclear ncRNAs, essential portions of the macronuclear somatic genome might be erroneously removed. Conversely, over-production of macronuclear ncRNA or under-production of scnRNAs might prevent the elimination of portions of the germline genome, which could be deleterious to the progeny. In support of this idea, cases of erroneous excision of segments of coding sequences have been observed. Short TA indels with all the characteristics of IESs that are nonetheless retained in a few copies in the macronuclear genome

have also been described and might represent failure to excise an IES (Duret et al., 2008).

Barring mating incompatibility, variability in germline-limited sequences and/or their excision will generate macronuclear genome variability that will be subject to natural selection. By epigenetic inheritance mechanisms, the selected macronuclear genome could influence rearrangement of the progeny genome, thus essentially passing on traits in a non-Mendelian fashion. Studies in both *Paramecium* and *Tetrahymena* have shown that the pattern of alternative outcomes of epigenetic inheritance can be influenced by environmental conditions (Arslanyolu and Doerder, 2000; Sonneborn, 1977; Sonneborn and Schneller, 1979). Also, as noted above, induced PTGS of *Paramecium* genes frequently results in their deletion from the macronucleus of progeny, through disruption of scnRNA selection processes. Ciliates thus present an attractive model system for studying the cross-generational ‘inheritance of acquired traits’, a subject that has come under increasing study in other organisms as well (Koonin and Wolf, 2009).

We have much to learn about ciliate genome rearrangement, including the nature of whole genome transcription of the germline micronucleus and somatic macronucleus, selection of scnRNAs and targeting of excision machinery. Better understanding of these events will shed light on the range of activities dependent on RNA-mediated epigenetic regulation in other eukaryotes. Besides further mechanistic studies, we eagerly anticipate full analyses of the micronuclear genomes of both *Paramecium tetraurelia* and *Tetrahymena thermophila*, which are underway. Comparative micronuclear and macronuclear genomic analyses of additional related species will also be instrumental in determining how genome rearrangement events are programmed and how they, in turn, shape the genome.

Funding

Work in our laboratory is supported by Centre National de la Recherche Scientifique (CNRS) (ATIP/AVENIR grant), Comité de Paris de la Ligue Nationale contre le Cancer (RS10/75-91), Agence Nationale de la Recherche (grant 2010-BLAN_1603-02), Ministère de l’Enseignement Supérieur et de la Recherche (PhD fellowship to M.L.A.) and Ville de Paris (‘Research in Paris’ program to R.S.C.).

Conflict of interest statement

The authors have declared no conflict of interest.

References

*Articles of special interest.

- Arnaiz, O., Gout, J.F., Betermier, M., Bouhouche, K., Cohen, J., Duret, L., Kapusta, A., Meyer, E. and Sperling, L. (2010) Gene expression in a paleopolyploid: A transcriptome resource for the ciliate *Paramecium tetraurelia*. *BMC Genomics* **11**, 547
- Aronica, L., Bednenko, J., Noto, T., DeSouza, L.V., Siu, K.W., Loidl, J., Pearlman, R.E., Gorovsky, M.A. and Mochizuki, K. (2008) Study of an RNA helicase implicates small RNA–noncoding RNA interactions in programmed DNA elimination in *Tetrahymena*. *Genes Dev.* **22**, 2228–2241
- Arslanyolu, M. and Doerder, F.P. (2000) Genetic and environmental factors affecting mating type frequency in natural isolates of *Tetrahymena thermophila*. *J. Eukaryot. Microbiol.* **47**, 412–418
- Aury, J.M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Camara, F., Dharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A.M., Kissmehl, R., Klotz, C., Koll, F., Le Mouel, A., Lepere, G., Malinsky, S., Nowacki, M., Nowak, J.K., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Betermier, M., Weissenbach, J., Scarpelli, C., Schachter, V., Sperling, L., Meyer, E., Cohen, J. and Wincker, P. (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**, 171–178
- *Baudry, C., Malinsky, S., Restituto, M., Kapusta, A., Rosa, S., Meyer, E. and Betermier, M. (2009) PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev.* **23**, 2478–2483
- Bednenko, J., Noto, T., DeSouza, L.V., Siu, K.W., Pearlman, R.E., Mochizuki, K. and Gorovsky, M.A. (2009) Two GW repeat proteins interact with *Tetrahymena thermophila* argonaute and promote genome rearrangement. *Mol. Cell Biol.* **29**, 5020–5030
- Berger, J.D. and Schmidt, H.J. (1978) Regulation of macronuclear DNA content in *Paramecium tetraurelia*. *J. Cell Biol.* **76**, 116–126
- Betermier, M. (2004) Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate *Paramecium*. *Res. Microbiol.* **155**, 399–408
- Blackburn, E.H., Pan, W.C. and Johnson, C.C. (1983) Methylation of ribosomal RNA genes in the macronucleus of *Tetrahymena thermophila*. *Nucleic Acids Res.* **11**, 5131–5145
- Bouhouche, K., Gout, J.F., Kapusta, A., Betermier, M. and Meyer, E. (2011) Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res.* **39**, 4249–4264
- Bourc’his, D. and Bestor, T.H. (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96–99
- Bourc’his, D. and Voinnet, O. (2010) A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science* **330**, 617–622
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R. and Hannon, G.J. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103
- Buhler, M. and Moazed, D. (2007) Transcription and RNAi in heterochromatic gene silencing. *Nat. Struct. Mol. Biol.* **14**, 1041–1048

- Buhler, M., Verdel, A. and Moazed, D. (2006) Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**, 873–886
- Cech, T.R. (1993) The efficiency and versatility of catalytic RNA: Implications for an RNA world. *Gene* **135**, 33–36
- Cervantes, M.D., Xi, X., Vermaak, D., Yao, M.C. and Malik, H.S. (2006) The CNA1 histone of the ciliate *Tetrahymena thermophila* is essential for chromosome segregation in the germline micronucleus. *Mol. Biol. Cell* **17**, 485–497
- Chalker, D.L., Fuller, P. and Yao, M.C. (2005) Communication between parental and developing genomes during tetrahymena nuclear differentiation is likely mediated by homologous RNAs. *Genetics* **169**, 149–160
- *Chalker, D.L. and Yao, M.C. (1996) Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of *Tetrahymena thermophila* with germ line-limited DNA. *Mol. Cell Biol.* **16**, 3658–3667
- Chalker, D.L. and Yao, M.C. (2001) Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in *Tetrahymena thermophila*. *Genes Dev.* **15**, 1287–1298
- Chalker, D.L. and Yao, M.C. (2011) DNA elimination in ciliates: transposon domestication and genome surveillance. *Annu. Rev. Genet.* **45**, 227–246
- *Cheng, C.Y., Vogt, A., Mochizuki, K. and Yao, M.C. (2010) A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in *Tetrahymena thermophila*. *Mol. Biol. Cell* **21**, 1753–1762
- Creamer, K.M. and Partridge, J.F. (2011) RITS-connecting transcription, RNA interference, and heterochromatin assembly in fission yeast. *Wiley Interdiscip. Rev. RNA* **2**, 632–646
- Cui, B. and Gorovsky, M.A. (2006) Centromeric histone H3 is essential for vegetative cell division and for DNA elimination during conjugation in *Tetrahymena thermophila*. *Mol. Cell Biol.* **26**, 4499–4510
- Cummings, D.J., Tait, A. and Goddard, J.M. (1974) Methylated bases in DNA from *Paramecium aurelia*. *Biochim. Biophys. Acta.* **374**, 1–11
- Doerder, F.P., Deak, J.C. and Lief, J.H. (1992) Rate of phenotypic assortment in *Tetrahymena thermophila*. *Dev. Genet.* **13**, 126–132
- Duharcourt, S., Butler, A. and Meyer, E. (1995) Epigenetic self-regulation of developmental excision of an internal eliminated sequence on *Paramecium tetraurelia*. *Genes Dev.* **9**, 2065–2077
- *Duharcourt, S., Keller, A.M. and Meyer, E. (1998) Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. *Mol. Cell Biol.* **18**, 7075–7085
- Duharcourt, S., Lepere, G. and Meyer, E. (2009) Developmental genome rearrangements in ciliates: A natural genomic subtraction mediated by non-coding transcripts. *Trends Genet.* **25**, 344–350
- Duret, L., Cohen, J., Jubin, C., Dessen, P., Gout, J.F., Mousset, S., Aury, J.M., Jaillon, O., Noel, B., Arnaz, O., Betermier, M., Wincker, P., Meyer, E. and Sperling, L. (2008) Analysis of sequence variability in the macronuclear DNA of *Paramecium tetraurelia*: A somatic view of the germline. *Genome Res.* **18**, 585–596
- Eisen, J.A., Coyne, R.S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J.R., Badger, J.H., Ren, Q., Amedeo, P., Jones, K.M., Tallon, L.J., Delcher, A.L., Salzberg, S.L., Silva, J.C., Haas, B.J., Majoros, W.H., Farzad, M., Carlton, J.M., Smith, R.K., Garg, J., Pearlman, R.E., Karrer, K.M., Sun, L., Manning, G., Elde, N.C., Turkewitz, A.P., Asai, D.J., Wilkes, D.E., Wang, Y., Cai, H., Collins, K., Stewart, B.A., Lee, S.R., Wilamowska, K., Weinberg, Z., Ruzzo, W.L., Wloga, D., Gaertig, J., Frankel, J., Tsao, C.C., Gorovsky, M.A., Keeling, P.J., Waller, R.F., Patron, N.J., Cherry, J.M., Stover, N.A., Krieger, C.J., Del Toro, C., Ryder, H.F., Williamson, S.C., Barbeau, R.A., Hamilton, E.P. and Orias, E. (2006) Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol.* **4**, e286
- El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R. and Lagrange, T. (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* **21**, 2539–2544
- Epstein, L.M. and Forney, J.D. (1984) Mendelian and non-Mendelian mutations affecting surface antigen expression in *Paramecium tetraurelia*. *Mol. Cell Biol.* **4**, 1583–1590
- Fan, Q. and Yao, M. (1996) New telomere formation coupled with site-specific chromosome breakage in *Tetrahymena thermophila*. *Mol. Cell Biol.* **16**, 1267–1274
- Fass, J., Joshi, N., Couvillion, M.T., Tonkin, L.A., Bowen, J., Gorovsky, M.A., Hamilton, E.P., Orias, E., Hong, K., Coyne, R.S., Eisen, J.A., Chalker, D.L., Lin, D. and Collins, K. (2011) Characterization of *Tetrahymena thermophila* macronuclear genome sites of DNA elimination by micronuclear genome sequencing. *G3: Genes, Genomes, Genetics* **1**, 515–522
- Forney, J.D. and Blackburn, E.H. (1988) Developmentally controlled telomere addition in wild-type and mutant *paramecia*. *Mol. Cell Biol.* **8**, 251–258
- Garnier, O., Serrano, V., Duharcourt, S. and Meyer, E. (2004) RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. *Mol. Cell Biol.* **24**, 7370–7379
- Girard, A. and Hannon, G.J. (2008) Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol.* **18**, 136–148
- Gratias, A. and Betermier, M. (2001) Developmentally programmed excision of internal DNA sequences in *Paramecium aurelia*. *Biochimie.* **83**, 1009–1022
- *Gratias, A. and Betermier, M. (2003) Processing of double-strand breaks is involved in the precise excision of *paramecium* internal eliminated sequences. *Mol. Cell Biol.* **23**, 7152–7162
- Gratias, A., Lepere, G., Garnier, O., Rosa, S., Duharcourt, S., Malinsky, S., Meyer, E. and Betermier, M. (2008) Developmentally programmed DNA splicing in *Paramecium* reveals short-distance crosstalk between DNA cleavage sites. *Nucleic Acids Res.* **36**, 3244–3251
- Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H. and Siomi, M.C. (2007) A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587–1590
- Harrison, G.S. and Karrer, K.M. (1985) DNA synthesis, methylation and degradation during conjugation in *Tetrahymena thermophila*. *Nucleic Acids Res.* **13**, 73–87
- Huvos, P. (2004a) A member of a repeat family is the source of an insertion–deletion polymorphism inside a developmentally eliminated sequence of *Tetrahymena thermophila*. *J. Mol. Biol.* **336**, 1061–1073
- Huvos, P. (2004b) Modular structure in developmentally eliminated DNA in *Tetrahymena* may be a consequence of frequent insertions and deletions. *J. Mol. Biol.* **336**, 1075–1086
- Jonsson, F., Postberg, J. and Lipps, H.J. (2009) The unusual way to make a genetically active nucleus. *DNA Cell Biol.* **28**, 71–78
- Joshua-Tor, L. and Hannon, G.J. (2011) Ancestral roles of small RNAs: An Ago-centric perspective. *Cold Spring Harb. Perspect. Biol.* **3**, a003772
- Kapitonov, V.V. and Jurka, J. (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.* **3**, e181
- *Kapusta, A., Matsuda, A., Marmignou, A., Ku, M., Silve, A., Meyer, E., Forney, J.D., Malinsky, S. and Betermier, M. (2011) Highly precise and developmentally programmed genome assembly in *Paramecium* requires ligase IV-dependent end joining. *PLoS Genet.* **7**, e1002049

- Karagianni, P., Amazit, L., Qin, J. and Wong, J. (2008) ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. *Mol. Cell Biol.* **28**, 705–717
- Karginov, F.V. and Hannon, G.J. (2010) The CRISPR system: Small RNA-guided defense in bacteria and archaea. *Mol. Cell* **37**, 7–19
- Karrer, K.M. and VanNuland, T.A. (2002) Methylation of adenine in the nuclear DNA of *Tetrahymena* is internucleosomal and independent of histone H1. *Nucleic Acids Res.* **30**, 1364–1370
- *Klobutcher, L.A. and Herrick, G. (1995) Consensus inverted terminal repeat sequence of *Paramecium* IESs: Resemblance to termini of Tc1-related and *Euplotes* Tec transposons. *Nucleic Acids Res.* **23**, 2006–2013
- Klobutcher, L.A. and Herrick, G. (1997) Developmental genome reorganization in ciliated protozoa: The transposon link. *Prog. Nucleic Acid Res. Mol. Biol.* **56**, 1–62
- Koonin, E.V. and Wolf, Y.I. (2009) Is evolution Darwinian or/and Lamarckian? *Biol. Direct* **4**, 42
- Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., Hata, K., Li, E., Matsuda, Y., Kimura, T., Okabe, M., Sakaki, Y., Sasaki, H. and Nakano, T. (2008) DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* **22**, 908–917
- Kurth, H.M. and Mochizuki, K. (2009) 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in *Tetrahymena*. *RNA* **15**, 675–685
- Larson, D.D., Umthun, A.R. and Shaiu, W.L. (1991) Copy number control in the *Tetrahymena* macronuclear genome. *J. Protozool.* **38**, 258–263
- Le Mouel, A., Butler, A., Caron, F. and Meyer, E. (2003) Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in *paramecia*. *Eukaryot. Cell* **2**, 1076–1090
- Lee, S.R. and Collins, K. (2006) Two classes of endogenous small RNAs in *Tetrahymena thermophila*. *Genes Dev.* **20**, 28–33
- Lejeune, E., Bayne, E.H. and Allshire, R.C. (2011) On the connection between RNAi and heterochromatin at centromeres. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 275–283
- *Lepere, G., Betermier, M., Meyer, E. and Duharcourt, S. (2008) Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev.* **22**, 1501–1512
- Lepere, G., Nowacki, M., Serrano, V., Gout, J.F., Guglielmi, G., Duharcourt, S., and Meyer, E. (2009) Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res.* **37**, 903–915
- Li, L. and Liu, Y. (2011) Diverse small non-coding RNAs in RNA interference pathways. *Methods Mol. Biol.* **764**, 169–182
- Lieber, M.R., Ma, Y., Pannicke, U. and Schwarz, K. (2004) The mechanism of vertebrate nonhomologous DNA end joining and its role in V(D)J recombination. *DNA Repair (Amst)* **3**, 817–826
- Liu, Y., Mochizuki, K. and Gorovsky, M.A. (2004) Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in *Tetrahymena*. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1679–1684
- Liu, Y., Song, X., Gorovsky, M.A. and Karrer, K.M. (2005) Elimination of foreign DNA during somatic differentiation in *Tetrahymena thermophila* shows position effect and is dosage dependent. *Eukaryot. Cell* **4**, 421–431
- *Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., Hunt, D.F. and Allis, C.D. (2007) RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in *Tetrahymena*. *Genes Dev.* **21**, 1530–1545
- Malone, C.D., Anderson, A.M., Motl, J.A., Rexer, C.H. and Chalker, D.L. (2005) Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of *Tetrahymena thermophila*. *Mol. Cell Biol.* **25**, 9151–9164
- Mansfield, R.E., Musselman, C.A., Kwan, A.H., Oliver, S.S., Garske, A.L., Davrazou, F., Denu, J.M., Kutateladze, T.G. and Mackay, J.P. (2011) Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *J. Biol. Chem.* **286**, 11779–11791
- Martindale, D.W., Allis, C.D. and Bruns, P.J. (1985) RNA and protein synthesis during meiotic prophase in *Tetrahymena thermophila*. *J. Protozool.* **32**, 644–649
- Matsuda, A., Shieh, A.W., Chalker, D.L. and Forney, J.D. (2010) The conjugation-specific Die5 protein is required for development of the somatic nucleus in both *Paramecium* and *Tetrahymena*. *Eukaryot. Cell* **9**, 1087–1099
- Matthews, A.G., Kuo, A.J., Ramon-Maiques, S., Han, S., Champagne, K.S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D.N., Walter, K.L., Utz, P.J., Shi, Y., Kutateladze, T.G., Yang, W., Gozani, O. and Oettinger, M.A. (2007) RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* **450**, 1106–1110
- Mayer, K.M. and Forney, J.D. (1999) A mutation in the flanking 5'-TA-3' dinucleotide prevents excision of an internal eliminated sequence from the *Paramecium tetraurelia* genome. *Genetics* **151**, 597–604
- Mitra, R., Fain-Thornton, J. and Craig, N.L. (2008) piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J.* **27**, 1097–1109
- *Mochizuki, K., Fine, N.A., Fujisawa, T. and Gorovsky, M.A. (2002) Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *tetrahymena*. *Cell* **110**, 689–699
- Mochizuki, K. and Gorovsky, M.A. (2004a) Conjugation-specific small RNAs in *Tetrahymena* have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev.* **18**, 2068–2073
- Mochizuki, K. and Gorovsky, M.A. (2004b) RNA polymerase II localizes in *Tetrahymena thermophila* meiotic micronuclei when micronuclear transcription associated with genome rearrangement occurs. *Eukaryot. Cell* **3**, 1233–1240
- Mochizuki, K. and Gorovsky, M.A. (2005) A Dicer-like protein in *Tetrahymena* has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes Dev.* **19**, 77–89
- Motl, J.A. and Chalker, D.L. (2011) Zygotic expression of the double-stranded RNA binding motif protein, Drb2p, is required for DNA elimination in the ciliate, *Tetrahymena thermophila*. *Eukaryot. Cell* **10**, 1648–1659
- Musselman, C.A., Mansfield, R.E., Garske, A.L., Davrazou, F., Kwan, A.H., Oliver, S.S., O'Leary, H., Denu, J.M., Mackay, J.P. and Kutateladze, T.G. (2009) Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem J.* **423**, 179–187
- Nikiforov, M.A., Smothers, J.F., Gorovsky, M.A. and Allis, C.D. (1999) Excision of micronuclear-specific DNA requires parental expression of pdd2p and occurs independently from DNA replication in *Tetrahymena thermophila*. *Genes Dev.* **13**, 2852–2862
- Noto, T., Kurth, H.M., Kataoka, K., Aronica, L., DeSouza, L.V., Siu, K.W., Pearlman, R.E., Gorovsky, M.A. and Mochizuki, K. (2010) The *Tetrahymena* argonaute-binding protein Giw1p directs a mature argonaute-siRNA complex to the nucleus. *Cell* **140**, 692–703
- Nowacki, M., Zagorski-Ostojka, W. and Meyer, E. (2005) Nowa1p and Nowa2p: Novel putative RNA binding proteins involved in trans-nuclear crosstalk in *Paramecium tetraurelia*. *Curr. Biol.* **15**, 1616–1628
- Nowak, J.K., Gromadka, R., Juszczuk, M., Jerka-Dziadosz, M., Maliszewska, K., Mucchielli, M.H., Gout, J.F., Arnaiz, O., Agier, N., Tang, T., Aggerbeck, L.P., Cohen, J., Delacroix, H., Sperling, L., Herbert, C.J., Zagulski, M. and Betermier, M. (2011) A functional

- study of genes essential for autogamy and nuclear reorganization in *Paramecium*. *Eukaryot. Cell* **10**, 363–372
- Orias, E. and Flacks, M. (1975) Macronuclear genetics of *Tetrahymena*. I. Random distribution of macronuclear genecopies in *T. pyriformis*, syngen 1. *Genetics* **79**, 187–206
- Plasterk, R.H., Izsvak, Z. and Ivics, Z. (1999) Resident aliens: The Tc1/mariner superfamily of transposable elements. *Trends Genet.* **15**, 326–332
- Rigal, M. and Mathieu, O. (2011) A “mille-feuille” of silencing: Epigenetic control of transposable elements. *Biochim. Biophys. Acta* **1809**, 452–458
- Ruiz, F., Krzywicka, A., Klotz, C., Keller, A., Cohen, J., Koll, F., Balavoine, G. and Beisson, J. (2000) The SM19 gene, required for duplication of basal bodies in *Paramecium*, encodes a novel tubulin, eta-tubulin. *Curr. Biol.* **10**, 1451–1454
- Sarkar, A., Sim, C., Hong, Y.S., Hogan, J.R., Fraser, M.J., Robertson, H.M. and Collins, F.H. (2003) Molecular evolutionary analysis of the widespread piggyBac transposon family and related “domesticated” sequences. *Mol. Genet. Genomics* **270**, 173–180
- Saveliev, S.V. and Cox, M.M. (1996) Developmentally programmed DNA deletion in *Tetrahymena thermophila* by a transposition-like reaction pathway. *EMBO J* **15**, 2858–2869
- Schoeberl, U.E. and Mochizuki, K. (2011) Keeping the soma free of transposons: Programmed DNA elimination in ciliates. *J. Biol. Chem.* **286**, 37045–37052
- Siomi, H. and Siomi, M.C. (2008) Interactions between transposable elements and Argonautes have (probably) been shaping the *Drosophila* genome throughout evolution. *Curr. Opin. Genet. Dev.* **18**, 181–187
- Siomi, M.C., Sato, K., Pezic, D. and Aravin, A.A. (2011) PIWI-interacting small RNAs: The vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12**, 246–258
- Smothers, J.F., Madiredi, M.T., Warner, F.D. and Allis, C.D. (1997) Programmed DNA degradation and nucleolar biogenesis occur in distinct organelles during macronuclear development in *Tetrahymena*. *J. Eukaryot. Microbiol.* **44**, 79–88
- Sonneborn, T.M. (1977) Genetics of cellular differentiation: Stable nuclear differentiation in eucaryotic unicells. *Annu. Rev. Genet.* **11**, 349–367
- Sonneborn, T.M. and Schneller, M.V. (1979) A genetic system for alternative stable characteristics in genomically identical homozygous clones. *Dev. Genet.* **1**, 21–46
- Stargell, L.A. and Gorovsky, M.A. (1994) TATA-binding protein and nuclear differentiation in *Tetrahymena thermophila*. *Mol. Cell Biol.* **14**, 723–734
- Sugai, T. and Hiwatashi, K. (1974) Cytologic and autoradiographic studies of the micronucleus at meiotic prophase in *Tetrahymena pyriformis*. *J. Protozool.* **21**, 542–548
- *Taverna, S.D., Coyne, R.S. and Allis, C.D. (2002) Methylation of histone h3 at lysine 9 targets programmed DNA elimination in *tetrahymena*. *Cell* **110**, 701–711
- Till, S., Lejeune, E., Thermann, R., Bortfeld, M., Hothorn, M., Enderle, D., Heinrich, C., Hentze, M.W. and Ladurner, A.G. (2007) A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. *Nat. Struct. Mol. Biol.* **14**, 897–903
- Volpe, T. and Martienssen, R.A. (2011) RNA interference and heterochromatin assembly. *Cold Spring Harb. Perspect. Biol.* **3**, a003731
- Watanabe, T., Tomizawa, S., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S., Iida, N., Hoki, Y., Murphy, P.J., Toyoda, A., Gotoh, K., Hiura, H., Arima, T., Fujiyama, A., Sado, T., Shibata, T., Nakano, T., Lin, H., Ichianagi, K., Soloway, P.D. and Sasaki, H. (2011) Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse *Rasgrf1* locus. *Science* **332**, 848–852
- Wion, D. and Casadesus, J. (2006) N6-methyl-adenine: An epigenetic signal for DNA–protein interactions. *Nat. Rev. Microbiol.* **4**, 183–192
- Wuitschick, J.D., Gershan, J.A., Lochowicz, A.J., Li, S. and Karrer, K.M. (2002) A novel family of mobile genetic elements is limited to the germline genome in *Tetrahymena thermophila*. *Nucleic Acids Res.* **30**, 2524–2537
- Yao, M.C., Choi, J., Yokoyama, S., Austerberry, C.F. and Yao, C.H. (1984) DNA elimination in *Tetrahymena*: A developmental process involving extensive breakage and rejoining of DNA at defined sites. *Cell* **36**, 433–440
- *Yao, M.C., Fuller, P. and Xi, X. (2003) Programmed DNA deletion as an RNA-guided system of genome defense. *Science* **300**, 1581–1584
- Yao, M.C. and Gorovsky, M.A. (1974) Comparison of the sequences of macro- and micronuclear DNA of *Tetrahymena pyriformis*. *Chromosoma* **48**, 1–18
- Yao, M.C., Yao, C.H. and Monks, B. (1990) The controlling sequence for site-specific chromosome breakage in *Tetrahymena*. *Cell* **63**, 763–772
- Zhang, H. and Zhu, J.K. (2011) RNA-directed DNA methylation. *Curr. Opin. Plant Biol.* **14**, 142–147
- Zufall, R.A., McGrath, C.L., Muse, S.V. and Katz, L.A. (2006) Genome architecture drives protein evolution in ciliates. *Mol. Biol. Evol.* **23**, 1681–1687

Received: 2 December 2011; Accepted: 15 February 2012; Accepted article online: 21 February 2012

